

https://dspace.org/

# A Search for Exoplanet Candidates in TESS 2 minute Light Curves Using Joint Bayesian Detection

**Jamila S. Taaki; Athol Kemball; Farzad Kamalabadi**

IOP PUBLISHING LTD
https://hdl.handle.net/10539/46231
WOS

Downloaded from a DSpace repository. https://dspace.org/

The Astronomical Journal, 170:14 (19pp), 2025 July https://doi.org/10.3847/1538-3881/add15a
© 2025. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

# A Search for Exoplanet Candidates in TESS 2 minute Light Curves Using Joint Bayesian Detection

Jamila S. Taaki[1], Athol J. Kemball[2,3], and Farzad Kamalabadi[4]
[1] Michigan Institute for Data Science, University of Michigan, 500 Church Road, Suite 600, Ann Arbor, Michigan 48109, USA; xiaziyna@gmail.com
[2] Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801-3074, USA
[3] Wits Centre for Astrophysics, School of Physics, University of the Witwatersrand, P.O. Box Wits, Johannesburg 2050, South Africa
[4] Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 306 N. Wright Street MC 702, Urbana, IL 61801-2918, USA
*Received 2024 July 16; revised 2025 April 24; accepted 2025 April 25; published 2025 June 5*

## Abstract

In this work, we apply an exploratory joint Bayesian transit detector, previously evaluated using Kepler data, to the 2 minutes simple aperture photometry light-curve data in the continuous viewing zone for the Transiting Exoplanet Survey Satellite (TESS) over 3 yr of observation. The detector uses Bayesian priors, adaptively estimated, to model unknown systematic noise and stellar variability incorporated in a Neyman–Pearson likelihood ratio test for a candidate transit signal; a primary goal of the algorithm is to reduce overfitting. The detector was adapted to the TESS data and refined to improve outlier rejection and suppress FA detections in postprocessing. The statistical performance of the detector was evaluated using transit injection tests, where the joint Bayesian detector achieves an 80.0% detection rate and a 19.1% quasi-false-alarm rate at a detection threshold $\tau = 10$; this is a marginal, although not statistically significant, improvement of 0.2% over a reference sequential detrending and detection algorithm. In addition, a full search of the input TESS data was performed to evaluate the recovery rate of known TESS Objects of Interest (TOIs) and to perform an independent search for new exoplanet candidates. The joint detector has a 73% recall rate and a 63% detection rate for known TOIs; the former considers a match against all detection statistics above threshold, while the latter considers only the maximum detection statistic.

*Unified Astronomy Thesaurus concepts:* Exoplanets (498); Transit photometry (1709); Bayesian statistics (1900); Computational methods (1965); Transit instruments (1708); Astronomy data analysis (1858); Likelihood ratio test (1942); Light curves (918); Astrostatistics (1882); Photometry (1234)

## 1. Introduction

The Transiting Exoplanet Survey Satellite (TESS; G. R. Ricker et al. 2014a) was launched in 2018 to survey $\sim 10^6$ nearby stars for planets.[5] TESS is the successor to two decades of wide-field exoplanet transit surveys, primarily CoRoT (M. Auvergne et al. 2009)[6] and Kepler (W. J. Borucki et al. 2010).[7] Such surveys are scientifically critical to understanding the population statistics of exoplanets by type and exoplanetary configuration (W. Zhu & S. Dong 2021). Innovations in statistical inference and specialized data processing have been critical in achieving the photometric precision (H. J. Deeg & R. Alonso 2018) required to robustly detect exoplanet transits in these data-intensive surveys (M. Jara-Maldonado et al. 2020). In particular, a number of specialized approaches have been developed to remove residual instrumental systematics that may otherwise mask astrophysical transit signals in light curves (J. D. Twicken et al. 2010; J. C. Smith et al. 2012; M. C. Stumpe et al. 2012, 2014; S. Roberts et al. 2013; D. Foreman-Mackey et al. 2015; S. Aigrain et al. 2016). In this work, we adapt a prior algorithm for joint Bayesian transit detection and systematic noise characterization (J. S. Taaki et al. 2020) that we previously evaluated on Kepler data, and apply it to TESS 2 minutes simple aperture photometry (SAP) light curves. The goals of this paper are to explore the modifications required and the numerical performance of this algorithm when applied to data with different residual instrumental systematics, and to independently search for new candidate exoplanets in the TESS data.

The TESS and Kepler missions differ significantly in their observing strategy and cadence. TESS has an instantaneous field of view (FOV) of $24^\circ \times 96^\circ$ spanned by four individual cameras, each containing four 2048 × 2048 pixel CCDs in their focal planes (G. R. Ricker et al. 2016). Individual pointings align the primary axis of the FOV along ecliptic longitude spanning a range of ecliptic latitudes from $\sim 6^\circ$ to the ecliptic pole (G. R. Ricker et al. 2016). Each pointing of the FOV is of ∼27 days duration, and is known as a TESS sector (G. R. Ricker et al. 2016). Each observing year in the primary mission spans one ecliptic hemisphere over 13 sectors equispaced in ecliptic longitude, and alternates to the other hemisphere the following year (G. R. Ricker et al. 2016). Approximately 2% of targets are continuously observed for the full year in the overlapping polar region known as the continuous viewing zone (CVZ; T. Barclay et al. 2018; N. L. Eisner et al. 2020). TESS produces Full-Frame Images (or FFIs) with a 30 minutes cadence, which are analyzed by the Quick-Look-Pipeline (QLP; C. X. Huang et al. 2020). The TESS Science Processing Operations Center (SPOC) pipeline (J. M. Jenkins et al. 2016; J. Twicken et al. 2019) also produces SAP light curves with 2 minutes cadence for a list of

[5] tess.mit.edu
[6] sci.esa.int/corot
[7] keplerscience.arc.nasa.gov

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

targets defined in the TESS Input Catalog (TIC; K. G. Stassun et al. 2018).[8] To date, there have been 7525 TESS Objects of Interest (TOIs) identified, 2262 of which are from the SPOC pipeline and a further 728 have been promoted from Community TESS Objects of Interest (CTOIs) to TOIs. Currently, 620 TOIs have been confirmed as exoplanets.[9]

In this work, we apply our prior Bayesian detection method to the 2 minutes SAP light curves (J. D. Twicken et al. 2010; J. M. Jenkins et al. 2016; R. L. Morris et al. 2020) obtained for the TESS CVZ targets over the first 3 yr of observation. Each year of data is searched independently, and where a target is observed in multiple years light-curve data were not combined. A total of 5535 year-long light curves were searched for exoplanets with orbital periods in the range 1–100 days. Our search is complementary to the SPOC search for exoplanet candidates and other searches of TESS 2 minutes SAP data. These include the Planet-Hunters-TESS project (N. L. Eisner et al. 2020), which identified 90 promising candidates in the first 2 yr of TESS operations using a combination of visual inspection and semiautomated vetting. The Weird detector (A. Wheeler & D. Kipping 2019; J. Chakraborty et al. 2020) was used to search the first year of TESS data for any interesting periodic phenomena using a technique based on phase dispersion minimization, and reported 28 planet-like signals. Further, I. Wong et al. (2020) performed a phase-curve study of 22 known transiting systems with light curves from the first year to obtain secondary-eclipse brightness modulation measurements, as well as other transit properties. D. L. Holdsworth et al. (2021) performed a search and identified 12 rapidly oscillating stars. We note the implementation of a systematics-insensitive periodogram (R. Angus et al. 2016; C. Hedges et al. 2020), a Python tool to search TESS targets from the CVZ for long-period signals and identify stellar rotation rates. This method uses a noise model that incorporates systematics alongside sinusoid signal terms, which are jointly fitted to form a periodogram. TESS light curves have also been used to study stellar properties, including an eclipsing binary (EB) catalog (A. Prša et al. 2022),[10] stellar flare statistics (A. D. Feinstein et al. 2020), and asteroseismology (M. N. Lund et al. 2021).

Wide-field transit surveys retain unmodeled residual instrumental errors after initial calibration and processing that compromise their high target photometric sensitivity (H. J. Deeg & R. Alonso 2018). The instrumental origins are diverse (R. Vanderspek et al. 2018; J. M. Jenkins 2020), and the systematic noise in the light curves is generally correlated both temporally and spatially across the imaging sensor (J. S. Taaki et al. 2024). The removal of residual systematics (detrending) is known to risk overfitting and distorting transit features (J. L. Christiansen et al. 2013; D. Foreman-Mackey et al. 2015), particularly if the systematic and transit variability occur over similar timescales and are correlated. In addition, the residual systematics may mimic transit signatures, and consequently lower transit detection performance. As mitigation, joint transit detection has been explored by D. Foreman-Mackey et al. (2015), who derived non-Bayesian joint maximum-likelihood estimates of transit signals and a low-rank linear systematic noise model. G. Kovács et al. (2016) performed joint fitting on K2 data but found no advantage over a sequential approach, citing false alarms (FAs) that may occur due to underfitting of systematics. L. J. Garcia et al. (2024) introduced a Python tool for computing joint likelihoods but replaced a white-noise model (D. Foreman-Mackey et al. 2015) with a time-correlated noise model. In the context of detrended TESS pre-search data conditioned PDCSAP flux light curves, they found this method to improve the detection of transits in the presence of short-timescale stellar variability (<1 day). Our joint Bayesian detector (J. S. Taaki et al. 2020) is nonsequential and marginalizes over an empirical systematics prior and a stochastic noise prior, posing a binary hypothesis test for the light curve either containing a transit or not containing a transit. A likelihood ratio test universally maximizes the probability of detection at any rate of FA for a binary hypothesis model (P. Moulin & V. V. Veeravalli 2018). Our algorithm is summarized in Section 2.1; full details, including the novelty and context of the algorithm relative to other methods, can be found in the original paper (J. S. Taaki et al. 2020).

There are several major differences between the Kepler long-cadence data analyzed in our prior work and the TESS 2 minutes data which necessitate modifications of our algorithm implementation. Relative to Kepler, a high fraction (∼13%) of 2 minutes TESS flux density measurements are missing or flagged. Each TESS observational sector comprises two ∼13.7 days orbits around the Earth, and data collection ceases for several hours between each orbit to downlink the data collected during this time (G. R. Ricker et al. 2016). In addition to this periodic gap, data are flagged or missing due to a 2–6 days momentum dump cycle (R. Vanderspek et al. 2018; N. M. Guerrero et al. 2021). Although our algorithm supports nonuniform data and does not require gap filling or interpolation, the cadence of missing data presents an ambiguity in the detected exoplanet period, in common with other TESS transit searches (N. M. Guerrero et al. 2021). In addition, each year-long SAP light curve contains ∼200,000 data values and poses general computational challenges (J. Twicken et al. 2019). For our algorithm, we performed detection tests on $10^8$ candidate transit models with a 10 minutes step size in candidate period and epoch. This is a high-performance computing (HPC) challenge, and was implemented in optimized form on the Blue Waters petascale system.[11]

In addition to the cadence differences, TESS and Kepler have different instrumental systematics. TESS has a lower combined differential photometric precision (J. L. Christiansen et al. 2012; M. Battley et al. 2021), achieving below 230 ppm for a 10th-magnitude star over 1 hr (M. M. Fausnaugh et al. 2018). Comparably, Kepler achieved 30 ppm for a 12th-magnitude dwarf star over 6.5 hr (half the transit time of an Earth-analog; J. Christiansen & P. Machalek 2010; R. L. Gilliland et al. 2011). Consequently, the expected signal-to-noise ratio and detection power for transit signals is lower for TESS than would be expected in Kepler data. This lowered photometric precision arises from the highly elliptical 13.7 days TESS orbit. TESS data contain short-term instances of large flux variations (T. Nguyen et al. 2018), arising from high pointing jitter as well as quasi-periodic flux density ramps which are tied to the 2–6 days momentum dump cycles. During these cycles noisy reaction wheels amplify pointing jitter, blurring point-spread functions, and reduce apparent flux (G. R. Ricker et al. 2014b).

[8] tess.mit.edu/science/tess-input-catalog/
[9] tess.mit.edu/toi-releases/
[10] http://tessebs.villanova.edu/
[11] https://bluewaters.ncsa.illinois.edu/hardware-summary

After data downlink every 13.7 days, ramps occur due to thermal changes as a result of changing solar illumination during pointing. Additionally, scattered-light features from the Earth and Moon exacerbate the reduction in apparent flux. In instances where the field is crowded, increased blending as a result of reduced per-pixel resolution may also contribute to lowered precision (Y. Lu et al. 2020). Reduced precision as well as quasi-periodic systematics are known to affect transit event statistics and present as false positives (FPs) in the TESS SPOC planet search (M. M. Fausnaugh et al. 2018), the TESS QLP search (N. M. Guerrero et al. 2021), and as described by M. Kunimoto et al. (2023). We adapted our algorithm implementation to filter FAs using a robustness metric similar to S. Seader et al. (2013) in addition to several other postprocessing metrics.

We evaluated our detector performance using single-transit injection tests on a sample of 1000 TESS CVZ light curves using limb-darkened transit signals. At a detection threshold of $\tau = 10$ the joint detector achieved a high detection rate of ∼80.0% with a 19.1% quasi-false-alarm (quasi-FA) rate. Our goal was to explore the joint detection algorithm on data with different systematics, and we make no claim of optimality. The joint detector performs comparably to standard sequential detrending and detection, with a possible slight performance increment (0.2%). In our full search of CVZ light curves, approximately 28% of light curves have detections. We achieve a 63% detection rate for any TESS TOIs present in CVZ light curves. The detection rate is defined as the rate of match between the TOI orbital period and the detected orbital period of the top statistic above the detection threshold, allowing for a 2–3 multiple of either period due to ambiguity in the orbital periods as described in N. M. Guerrero et al. (2021). We achieve a 73% recall rate of TOIs, where the recall rate is defined as the rate of match between a TOI's orbital period and the best-matched orbital period above the detection threshold. After additional manual vetting to remove clear astrophysical FPs such as EBs, as well as instrumental and statistical sources of FA, we do not unambiguously identify any new candidate exoplanets.

The paper is organized as follows. Section 2 describes the joint detector and implementation for TESS data. In Section 3, we report the results from our injection and detection results applied to TESS data. Section 4 contains a discussion of recovered candidates, and conclusions are presented in Section 5.

## 2. Methods

In this work, we apply a joint Bayesian transit detector (Detector A; J. S. Taaki et al. 2020) to TESS CVZ SAP light-curve data. In this section, we describe the refinements and implementation of this algorithm as applied to the TESS data. A list of commonly used notation is provided in Appendix A.

### 2.1. Detection Model

In this section we provide a concise summary of Detector A, originally described by J. S. Taaki et al. (2020). We retain here the nomenclature of the original paper. After median subtraction and normalization, the light curve for star $i \in I$ (where $I$ is the set of integers) is denoted as the vector $\boldsymbol{y}_i$ of length $N$ time samples. The light-curve vector $\boldsymbol{y}_i$ is modeled as the vector sum of a systematic noise $\boldsymbol{l}_i$ term, a stochastic or statistical noise $\boldsymbol{s}_i$ term, and a possible transit signal $\boldsymbol{t} \in \boldsymbol{T}$ (where $\boldsymbol{T}$ is the parent set of transit signals). The term $\boldsymbol{s}_i$ includes both residual stellar variability, which is not modeled separately, and statistical error of a nonsystematic nature.

The transit detection framework is based on a binary hypothesis test: (i) $H_0$, the null hypothesis that $\boldsymbol{y}_i$ contains no transit signal; and (ii) $H_1$, the alternative hypothesis that $\boldsymbol{y}_i$ does indeed contain a transit signal $\boldsymbol{t}$ (J. S. Taaki et al. 2020; Equations (1) and (2)):

$$H_0\colon \boldsymbol{y}_i = \boldsymbol{s}_i + \boldsymbol{l}_i, \tag{1}$$

$$H_1\colon \boldsymbol{y}_i = \boldsymbol{t} + \boldsymbol{s}_i + \boldsymbol{l}_i. \tag{2}$$

The detection test can be stated as (J. S. Taaki et al. 2020; Equation (3))

$$T(\boldsymbol{y}_i) \equiv \mathcal{L}_i(\boldsymbol{y}_i) = \frac{p_1(\boldsymbol{y}_i)}{p_0(\boldsymbol{y}_i)} \underset{H_0}{\overset{H_1}{\gtrless}} \tau, \tag{3}$$

where $\mathcal{L}_i(\boldsymbol{y}_i)$ is a likelihood ratio test (S. M. Kay 1993; L. Wasserman 2013) which is Neyman–Pearson optimal, $p_h(\boldsymbol{y}_i) \equiv p(\boldsymbol{y}_i|H_h)$ is the conditional probability of observing $\boldsymbol{y}_i$ under hypothesis $H_h \in \{H_0, H_1\}$, and $\tau$ is the detection threshold.

This detector adopts a common low-rank linear model for the systematic noise contribution $\boldsymbol{l}_i$, expanded in terms of a set $\boldsymbol{v}_i$ of common basis vectors (CBVs) across the sensor (G. Kovács et al. 2005; M. C. Stumpe et al. 2012):

$$\boldsymbol{l}_i[n] = \sum_{k=1}^{K} c_k^i \boldsymbol{v}_k[n], \tag{4}$$

where $n$ denotes a discrete time sample index, and $K$ is the model rank ($K \ll N$; J. S. Taaki et al. 2020; Equation (4)). Each $c_k^i$ is a coefficient weighting of $\boldsymbol{v}_k$ for light curve $i$. We denote the vector of coefficients for a light curve as $\boldsymbol{c}_i = [c_1^i, c_2^i, ...c_k^i]^T$ and the set of vector coefficients for all light curves on the sensor as $\{\boldsymbol{c}_i\colon i \in I\}$.

The noise terms $\boldsymbol{s}_i$ and $\boldsymbol{l}_i$ are unknown a priori. This detector constrains the noise terms using empirical Gaussian priors $\boldsymbol{s}_i \backsim \mathcal{N}(\boldsymbol{0}, \mathrm{Cov}_{s,i})$ and $\boldsymbol{c}_i \backsim \mathcal{N}(\mu_{\boldsymbol{c}_i}, \mathrm{Cov}_{\boldsymbol{c}})$, where the coefficients $\boldsymbol{c}_i$ act as proxies for the systematic noise $\boldsymbol{l}_i$ (J. S. Taaki et al. 2020). The parameters of the models are obtained by empirical inference and, for the current application of the detector, this is described in Sections 2.2.4 and 2.2.3. The joint noise is distributed as $\boldsymbol{l}_i + \boldsymbol{s}_i \sim \mathcal{N}(\boldsymbol{V}\mu_{\boldsymbol{c}_i}, \mathrm{Cov}_{s,i} + \boldsymbol{V}\mathrm{Cov}_{\boldsymbol{c}}\boldsymbol{V}^{\boldsymbol{T}})$, where $V$ is an $N \times K$ matrix with $\boldsymbol{v}_k$ as column vectors (J. S. Taaki et al. 2020). Using shifted data $\hat{\boldsymbol{y}}_i = \boldsymbol{y}_i - \boldsymbol{V}\mu_{\boldsymbol{c}_i}$ (J. S. Taaki et al. 2020), a detector may be equivalently formed for a fixed signal $\boldsymbol{t}$ in the presence of joint noise $\boldsymbol{z}_i \backsim \mathcal{N}(\boldsymbol{0}, \mathrm{Cov}_{z,i})$, where the joint noise covariance is $\mathrm{Cov}_{z,i} = \mathrm{Cov}_{s,i} + \boldsymbol{V}\mathrm{Cov}_{\boldsymbol{c}}\boldsymbol{V}^{\boldsymbol{T}}$. When detecting a fixed signal in Gaussian noise, a matched filter is a Neyman–Pearson optimal detector with a test statistic $T_i(\hat{\boldsymbol{y}}_i)$ and equivalent to Equation (3) (J. S. Taaki et al. 2020) and first used in Equation (1) of J. M. Jenkins et al. (2002):

$$T_i(\hat{\boldsymbol{y}}_i) = \frac{\hat{\boldsymbol{y}}_i^T \mathrm{Cov}_{z,i}^{-1}\boldsymbol{t}}{\sqrt{\boldsymbol{t}^T \mathrm{Cov}_{z,i}{}^{-1}\boldsymbol{t}}} \underset{H_0}{\overset{H_1}{\gtrless}} \tau. \tag{5}$$

In this detector framework, exoplanet transits are searched for by computing transit detection tests $T_i(\hat{\boldsymbol{y}}_i)$ for a discretized set of candidate signals $\boldsymbol{t} \in \boldsymbol{T}$. The functional form of candidate transit signals, as well as the discretization of the search space in the current application, are described in Section 2.2.5.

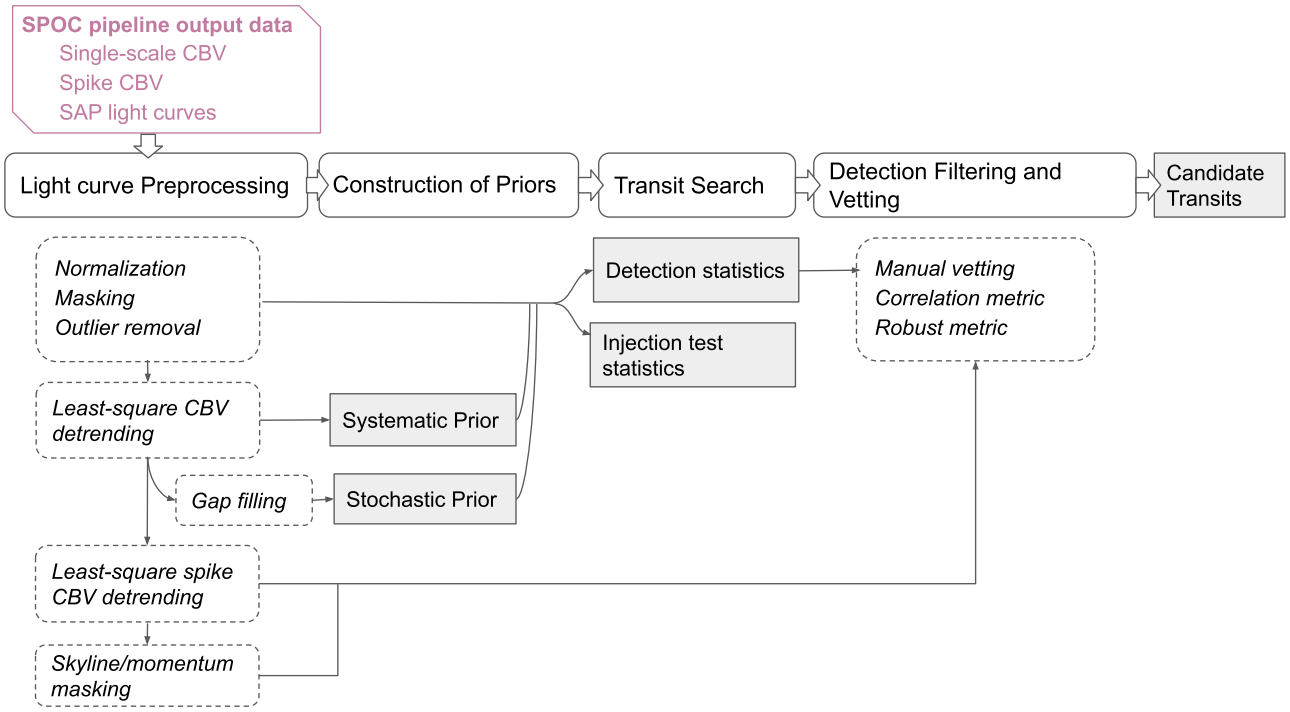



**Figure 1.** A schematic of the detector data processing pipeline. The input data set obtained from the SPOC pipeline output is shown in pink. Processing steps are shown in rounded boxes; major processing steps are shown with a solid outline with constituent component processing steps below with dashed outlines. Boxes indicating data output products are shaded in gray.

### 2.2. Implementation

The implementation of the data processing pipeline that applies this detector to the TESS data is shown in Figure 1. The flow of control moves from top to bottom and from left to right in the figure. The multiple stages and tunable parameters for the pipeline are described in this section. The nature of the TESS data requires implementation refinements relative to our previous application of the detector to Kepler data (J. S. Taaki et al. 2020). These changes relate to (i) the observing cadence, flagged data characteristics, and multisectoral nature of the TESS data; (ii) the computational optimization required to address the size of the data set; and, (iii) improved postprocessing and filtering of candidate detections.

The implementation is described as follows. Section 2.2.1 describes the input light-curve data set and Section 2.2.5 defines the candidate transit search space $\boldsymbol{T}$. The nature of missing and flagged TESS data, and their treatment, is described in Section 2.2.2. The choice of stochastic and systematic noise priors is described in Section 2.2.4 and Section 2.2.3. Post-detector processing is summarized in Section 2.2.6. The computational complexity of the detector and its efficient multisector computation is described in Appendix B.

#### 2.2.1. Target Light Curves and SPOC CBVs

As noted above, the input data set to our detector pipeline comprises the TESS 2 minutes SAP light curves (J. M. Jenkins et al. 2016) from the CVZ region for each of the first 3 yr of operations. These SAP light curves were obtained from the output of the TESS SPOC pipeline, along with the single-scale and spike CBVs provided by the SPOC analysis (see top left of Figure 1). The single-scale CBVs are not separated in timescale (J. C. Smith et al. 2012) in contradistinction to those formed using a multiscale wavelet basis (M. C. Stumpe et al. 2014); they are used here to form a light-curve systematics prior and as an orthogonal systematics basis (J. C. Smith et al. 2012; M. C. Stumpe et al. 2012). The spike CBVs contain large-amplitude narrow temporal spikes (P. Tenenbaum & J. M. Jenkins 2018; M. N. Lund et al. 2021); they are used here during detection filtering and vetting.

The CVZ has a high degree of data completeness and allows the detection of long-period signals. In years 1 and 3 TESS observed in the southern ecliptic hemisphere, and in year 2 in the northern ecliptic hemisphere. The number of targets per year was 1851 (year 1), 2046 (year 2), and 1922 (year 3). The total number of unique observed targets in our data set is 5535. We note that the TESS 2 minutes cadence targets varied between the primary (years 1 and 2) and extended mission (year 3) (M. Fausnaugh et al. 2021).

#### 2.2.2. Normalization, Masking, and Outlier Removal

As the first step in preprocessing (Figure 1) each input light curve $\boldsymbol{y}_{\rm in}$ was median-normalized as $\boldsymbol{y} = \boldsymbol{y}_{\rm in}/{\rm med}(\boldsymbol{y}_{\rm in}) - 1$, where ${\rm med}(\boldsymbol{y}_{\rm in})$ is the median of $\boldsymbol{y}_{\rm in}$. Nonnumeric (NaN) values were excluded in the normalization.

As noted above, TESS has a high rate of missing ($\sim$9%) and flagged ($\sim$4%) data compared to Kepler and at a significantly different cadence. Missing data arise due to a 2–6 days momentum dump cycle, a $\sim$13.7 days downlink cycle, and stray light from the Earth and Moon (R. Vanderspek et al. 2018; N. M. Guerrero et al. 2021). Figure 2 shows the average density of missing data in the year 1 input data set light curves over observing sector and time. Figure 3 shows the breakdown of quality-flagged data by systematic effect over year 1. We masked all missing and quality-flagged cadences (J. M. Jenkins et al. 2016; R. Vanderspek et al. 2018). Outliers were further removed in each light curve while retaining negative values by

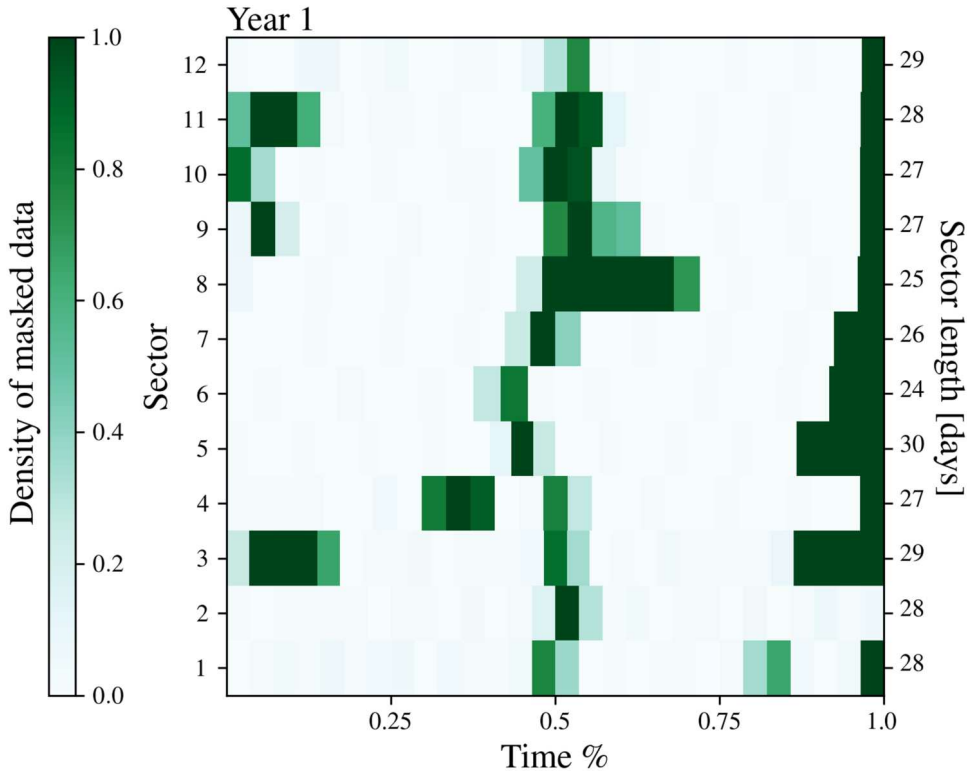



**Figure 2.** The fraction of missing data in the year 1 light curves, averaged per day and across all target light curves. Each row represents a sector of data. Sectors are variable length; the total number of days in a sector is displayed on the right axis. Each block represents the average fraction of data excluded (due to a quality flag or missing cadence) among the selected light curves, averaged over a day (approximately 720 cadences). Over each sector, most missing data occur at the midpoint due to a data downlink. However, missing data also occur throughout the remaining observational window.

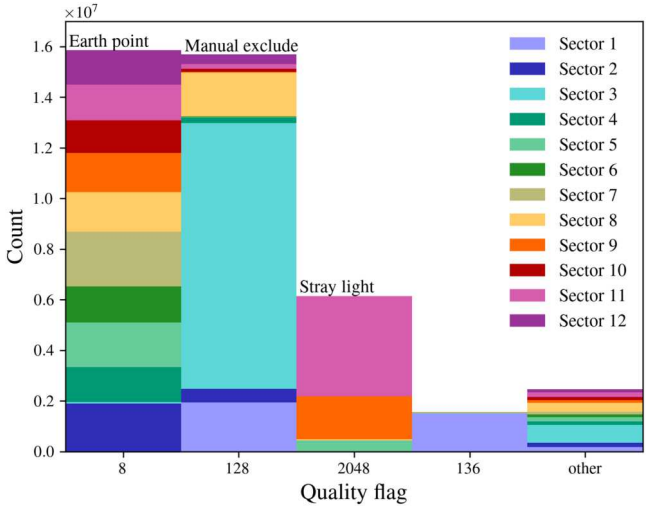



**Figure 3.** The frequency of quality-flagged cadences as they occur in year 1 light curves. The most commonly occurring quality flags have the following keys and descriptors: (8) spacecraft is in Earth point; (128) manual exclude due to an anomaly; (2048) stray light from Earth to Moon in camera FOV; (136) unspecified; and (other) all other quality flags. Broadly, the majority of flagged data is due to the spacecraft being in Earth point, and uniformly occurs over different sectors. We justify masking all flagged data as the three most commonly occurring flags are dominant and indicate a high likelihood of photometric error.

using $3\sigma$ positive iterative thresholding:

$$\boldsymbol{y}[n] = \begin{cases} \boldsymbol{y}_{\rm in}[n], & \text{if } \boldsymbol{y}_{\rm in}[n] \leqslant 3\sigma' \\ \boldsymbol{y}[n] \sim \mathcal{N}(0,\, \sigma'), & \boldsymbol{y}_{\rm in}[n] > 3\sigma' \end{cases}, \tag{6}$$

$\sigma' = \sigma(\boldsymbol{y}_{\rm in}|_{|\boldsymbol{y}_{\rm in}| \leqslant 3\sigma(\boldsymbol{y}_{\rm in})})$, where $\sigma(\boldsymbol{y})$ is the sample standard deviation of $\boldsymbol{y}$, and $\boldsymbol{y}|_{f(\boldsymbol{y})}$ denotes a subvector of $\boldsymbol{y}$ where the condition $f(\boldsymbol{y})$ is true. In our prior work searching Kepler data (J. S. Taaki et al. 2020) we did not perform outlier removal or data masking; these algorithm refinements were added here to improve performance further.

### 2.2.3. Detrending and Construction of Coefficient Systematic Prior

In earlier work (J. S. Taaki et al. 2020) the basis set $\{\boldsymbol{v}_k : k \in K\}$ was estimated using principal component analysis (PCA) on the collection of light curves. In this implementation, the TESS single-scale cotrending basis vectors (CBVs) were adopted directly as the basis set $\{\boldsymbol{v}_k : k \in K\}$ because the TESS CBVs are based on a similar technique to PCA, singular value decomposition, and have also undergone additional data filtering and analysis to minimize overfitting nonsystematic variability (J. C. Smith et al. 2012; M. C. Stumpe et al. 2012). TESS provides an independent set of 8 or 16 such CBVs for each sector, camera, and CCD set. An example set is shown in Figure 4. We note that systematic flux ramps occurring before and after data downlinks (midway at 13.7 days and at the end of a sector) are removed in the TESS analysis before the CBVs are formed. Accordingly, these flux ramps are not removed in our detrending analysis, the effect of which is described in further detail in Section 3.

A least-squares fit of the basis to a light curve (Equation (4)) without prior constraint may overfit and distort transits (J. C. Smith et al. 2012; M. C. Stumpe et al. 2012). As in J. S. Taaki et al. (2020), we formed a systematic prior $p(\boldsymbol{c}) \sim \mathcal{N}(\mu_{\boldsymbol{c}_i}, \mathrm{Cov}_{\boldsymbol{c}_i})$ on the coefficients $\boldsymbol{c}_i$ that is marginalized

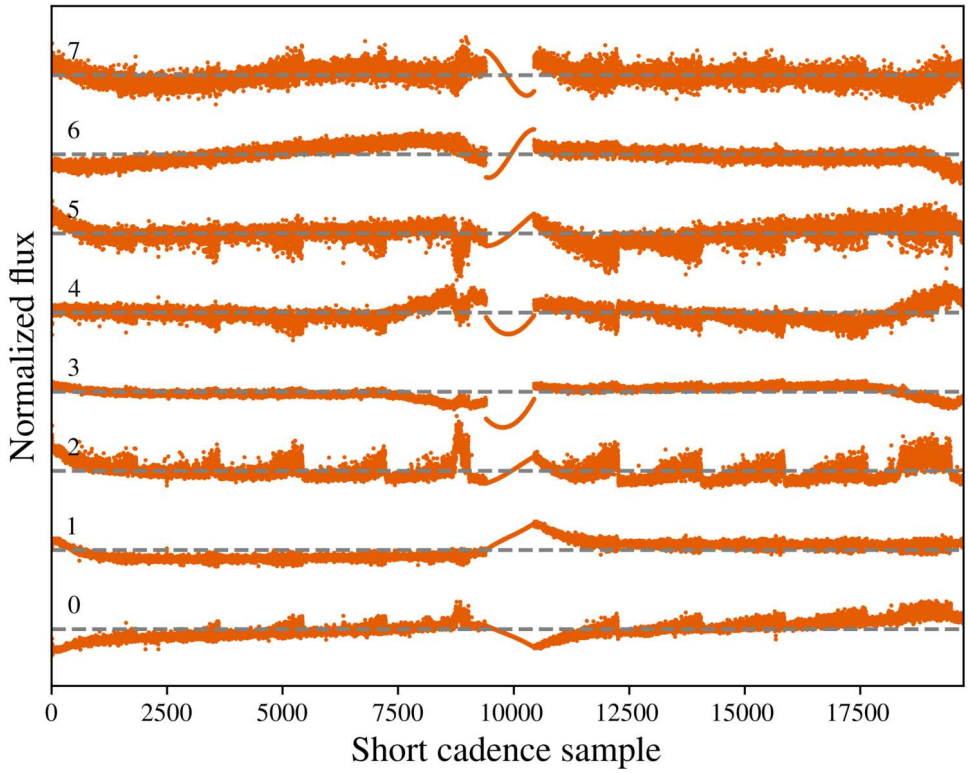



**Figure 4.** An example of 8/16 TESS cotrending basis vectors (CBVs) for sector 2, camera 4, and CCD 3 are shown at a 2 minutes cadence. The CBVs capture systematics induced by the TESS sensor, in particular the 2–6 days visible periodic pointing jitter. Ramps at the beginning of each 13.7 days orbit are likely due to thermal effects following the pointing slew during sector downlink. The sector downlink at ∼13.7 days appears in the middle of the plot. The data in this gap are masked in the current analysis.

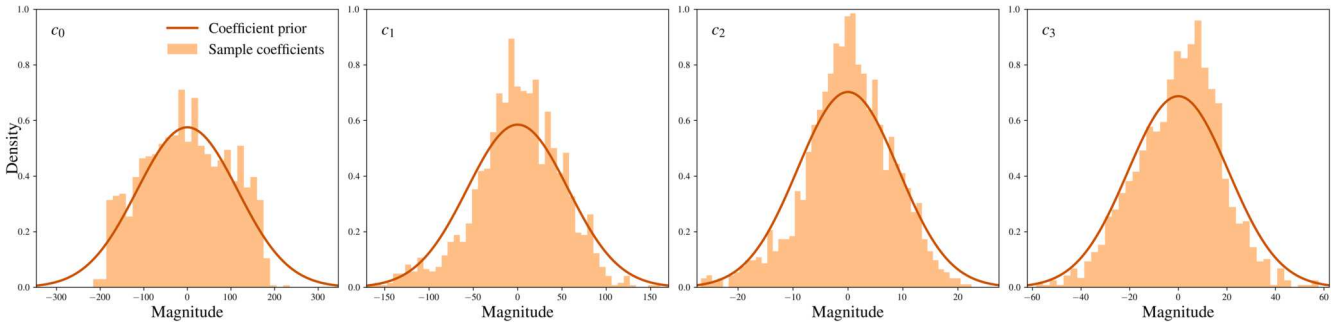



**Figure 5.** The density of empirical least-squares coefficients $\hat{\boldsymbol{c}}$ for the four leading basis vectors in sector 33, camera 4 and CCD 4, chosen as a random representative example. A fitted Gaussian prior is overlaid. The samples are heavy tailed in some cases, however they are reasonably close to the fitted Gaussian.

over in the detector and constrains overfitting. We formed a model of the covariance of coefficient vectors $\mathrm{Cov}_{\boldsymbol{c}}$ from the sample covariance from least-squares fits to all preprocessed light curves (including those outside the CVZ) for each combination of sector, camera, and CCD. As in J. S. Taaki et al. (2020), although each least-squares fit may be susceptible to overfitting, assuming independent astrophysical signals leads to a reasonable assumption of unbiased sample statistics.

The mean coefficient vector $\mu_{\boldsymbol{c}_i}$ was obtained from the least-squares fit for the light curve under consideration. In Figure 5, coefficient priors are shown alongside a histogram of fitted least-squares fits from all light curves among a chosen sector CCD and channel. A Gaussian prior is used here as a reasonable choice for the broad light-curve sample, as it inherently regularizes the data and easily captures correlations between systematic coefficients (J. S. Taaki et al. 2020). However, empirical priors modeled on physical parameters as in Presearch Data Conditioning Maximum A Posteriori (J. D. Twicken et al. 2010) may be more realistic, and thus the Gaussian prior may be underpowered over a subset of light curves. An empirical prior is not tractable for our Bayesian estimator.

### 2.2.4. Construction of the Stochastic Prior

The stochastic noise includes primary contributions from stellar variability, read noise, and Poisson noise (R. L. Gilliland et al. 2011), which can be reasonably assumed to be asymptotically Gaussian $\boldsymbol{s} \sim \mathcal{N}(0, \mathrm{Cov}_{\boldsymbol{s}})$ in combination. As in our prior work (J. S. Taaki et al. 2020), we assumed the stochastic noise is zero-mean, stationary (per sector), and completely described by the covariance $\mathrm{Cov}_{\boldsymbol{s}} \in \mathcal{R}^{N\times N}$. As in that work, we estimate $\mathrm{Cov}_{\boldsymbol{s},i}$ for light curve $i \in I$ in each sector using the spectrally smoothed autocorrelation computed from

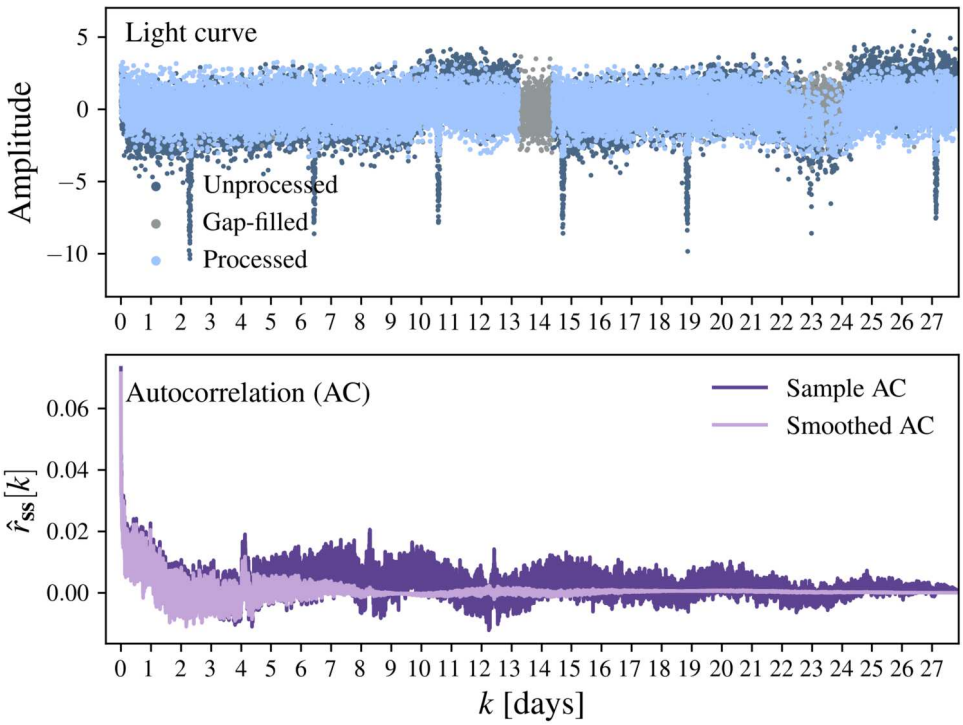



**Figure 6.** An example light curve (TID: 349518145, sector 1) with a transit signal is shown (top) with the associated fitted autocovariance $\mathrm{Cov}_{s,i}[s-t] = f(|s-t|) = f(k) = \hat{r}_{ss}(k)$ (below). In the top panel, the unprocessed light curve is shown in dark blue (background); this light curve is detrended and thresholded, shown in light blue, and gap-filled with random samples, shown in gray. The resulting light curve of light blue and gray points (foreground) is used to estimate the stochastic sample autocovariance (below, dark purple). The smoothed periodogram is shown in light purple. The smoothing reduces estimator variance and suppresses unfiltered transit variability, however it may also suppress stochastic variability.

least-squares detrended and $3\sigma$ thresholded light-curve data. Computing the estimated and smoothed spectrum requires uniform time sampling; gaps in the TESS light curves were filled with randomly selected portions of the light curve from the same sector. However, as shown in Figure 1, gap filling is only required for this spectral estimation of the stochastic prior; it is not required for computing the detection statistic (Equation (5)).

Estimates of $\mathrm{Cov}_{s,i}[t, u]$: $|t-u| = k$ for larger lags have fewer data samples and therefore greater estimation variance. The unsmoothed sample spectrum values, corresponding to the Fourier transform of the light-curve samples squared, are each independently $\chi^2_2$ distributed in the white-noise case (S. Kay 1999; J. T. VanderPlas 2018). The frequency binning of the spectral values averages independent $\chi^2_2$ variables, reducing the variance of spectral values. The estimation variance of autocorrelation samples, as the inverse Fourier transform of the smoothed periodogram samples, is thus reduced. This has the effect of suppressing the covariance at large lag values. We use a spectral smoothing window of 3 frequency bins. In Figure 6, an example of a light curve and the associated spectrally smoothed autocorrelation estimate is shown and compared to the sample autocorrelation, which has been suppressed at larger lags.

We note that this estimator is nonparametric and empirical. The spectral smoothing regularizes the stellar covariance model but may not entirely remove correlations due to transit eclipses. Physically motivated covariance models for stellar activity are described by R. Angus et al. (2017) and C. E. Rasmussen (2003).

Figure 7 shows a sample of light-curve noise models over sector 14 of the year 2 data. The relative $1\sigma$ contribution of the stochastic and systematic noise terms are shown over time. The noise models are highly dependent on the light curve under consideration.

#### 2.2.5. Transit Search: Detection Tests and Transit Search Space

A transit search was performed on each light curve $i \in \mathcal{I}$ to find any transit signal $\boldsymbol{t} \in \boldsymbol{T}$ with a detection statistic (Equation (5)) above the threshold $T_i(\hat{\boldsymbol{y}}_i) > \tau$. As in J. S. Taaki et al. (2020), we adopted a search space $\boldsymbol{T}$ of normalized box transit functions $\boldsymbol{b}$ parametrized by transit duration $d$, period $P$, and transit epoch $e$:

$$\boldsymbol{b}_{P,e,d}[n] = \begin{cases} 1, & \text{if } (n-e) \bmod P \leqslant d \\ 0, & \text{otherwise} \end{cases}. \tag{7}$$

The detector is invariant to the transit depth, as described in Appendix B, and this parameter was therefore not included in the search space.

As shown in Appendix C, for our matched filter detector the reduction in detection power using a box transit function of duration $d'$ instead of a true limb-darkened transit of duration $d$ (K. Mandel & E. Agol 2002) is given by $\frac{\mathbb{E}[T(\boldsymbol{y};\boldsymbol{b}_d)]}{\mathbb{E}[T(\boldsymbol{y};\boldsymbol{t}_d)]} = \boldsymbol{t}_d^T \boldsymbol{b}_{d'}$. When $d' = d$, we found in numerical simulation that $\frac{\mathbb{E}[T(\boldsymbol{y};\boldsymbol{b}_d)]}{\mathbb{E}[T(\boldsymbol{y};\boldsymbol{t}_d)]} \approx 0.92$. However, the optimal box that produces the minimal reduction in detection power may be obtained by a slightly narrower duration $d' < d$ in practice. Allowing $d'$ to vary among the discrete set of durations searched in Table 1 and simulating numerically with 1000 limb-darkened single-transit events as described in Section 2.3, we found $\frac{\mathbb{E}[T(\boldsymbol{y};\boldsymbol{b}_d)]}{\mathbb{E}[T(\boldsymbol{y};\boldsymbol{t}_d)]} \approx 0.96$.

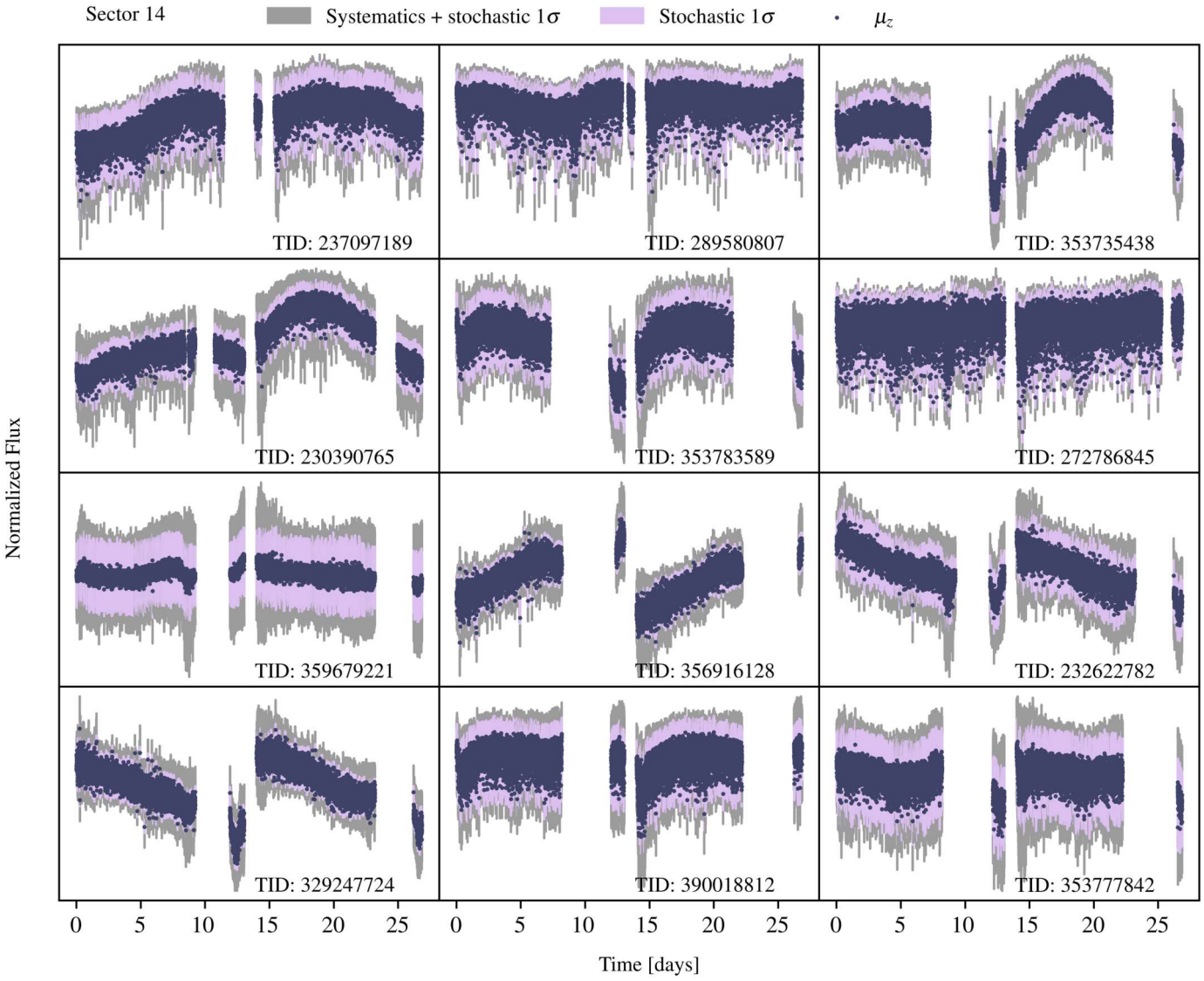



**Figure 7.** The $1\sigma$ stochastic and systematic noise levels over time are shown for a random selection of light curves from year 2, sector 14. The mean noise term $\mu_z$ (dark purple), the $1\sigma$ stochastic noise level $\sqrt{\mathrm{Cov}_s[n, n]}$ (light purple), and the $1\sigma$ joint stochastic and systemic noise level $\sqrt{\mathrm{Cov}_z[n, n]}$ (gray) are overplotted for each light curve. These noise covariance terms are described in Section 2.1. Although the relative contribution of the stochastic and systematic noise varies, they are generally of a similar magnitude. The masking is highly nonuniform between light curves in the same sector. We note that off-diagonal covariance terms are not depicted here; the full covariance of the light curve in the first panel (TID: 237097189) is shown in Figure 8.

We excluded all candidate transits with fewer than three transits present in the data, and required each transit to occur on a minimum of one nonmasked cadence. For each year of data, we evaluated approximately $|\boldsymbol{T}| \sim 2 \times 10^8$ candidate transit signals per light curve, where the transit search space $\boldsymbol{T}$ was parameterized as in Table 1. The step sizes in duration $d$, period $P$, and epoch $e$ were chosen to afford a high level of correlation between a real transit signal and the best-matched candidate signal (J. M. Jenkins et al. 1996; J. M. Jenkins et al. 2010). As candidate durations under 1 hr are possible (K. G. Stassun et al. 2019), and as further informed by the range of the Kepler Transiting Planet Search (J. M. Jenkins et al. 2010), our transit duration range spans 1–16 hr.

A slight mismatch in period $P$ may significantly affect detection statistics; a full exploration of template mismatch is described by J. M. Jenkins et al. (1996) and S. Seader et al. (2013). The maximum error in period $P$ is half the period step size: $\frac{\delta P}{2}$. After $M$ single-transit events, the mismatch in the position of the $M$th transit event will be $\frac{M\delta P}{2}$. Assuming a period $P$ with this worst-case period mismatch, the epoch $e$ which would minimize the worst-case single-transit offset, and therefore provide the closest signal match in a search, is that which provides the best match for (approximately) the middle transit event. In this case, the maximum offset of a single event is approximately the number of transits in half of the light curve $\frac{N}{2P}$ multiplied by the maximum error in period as $\frac{N}{2P} \cdot \frac{\delta P}{2} = \frac{N\delta P}{4P}$. Since the mismatch error is greater for shorter periods, we use a finer step size of $\delta P = 10$ minutes for periods $P \leqslant 27$ days and $\delta P = 20$ minutes for $P > 27$ days. As an example, a transit signal with period $P = 8$ days can incur an approximate worst-case error in the location of a single-transit event of $\sim$2 hr; if the transit duration is less than 2 hr, the template and transit event may be entirely mismatched. When calculating the matched filter test (Equation (5)), each candidate transit in a sector was allowed to shift $\pm$2 hr in

**Table 1**
Transit Search Space

| Transit Parameter | Range | Step Size |
|---|---|---|
| Orbital period | $P \in [1, 100]$ days | $\delta P \in \{10, 20\}$ minutes |
| Transit duration | $d \in \{1, 2, 3, 4, 6, 8, 10, 12, 14, 16\}$ hr | |
| Epoch | $e \in [0, P]$ | $\delta e \in \{10, 20\}$ minutes |

**Note.** The sample integration time is $\triangle t_{\rm LC} = 2$ minutes. The step size of period $\delta P$ and epoch $\delta e$ is 10 minutes for candidate periods less than the length of a sector ~27 days, otherwise $\delta P$ and $\delta e$ are 20 minutes.

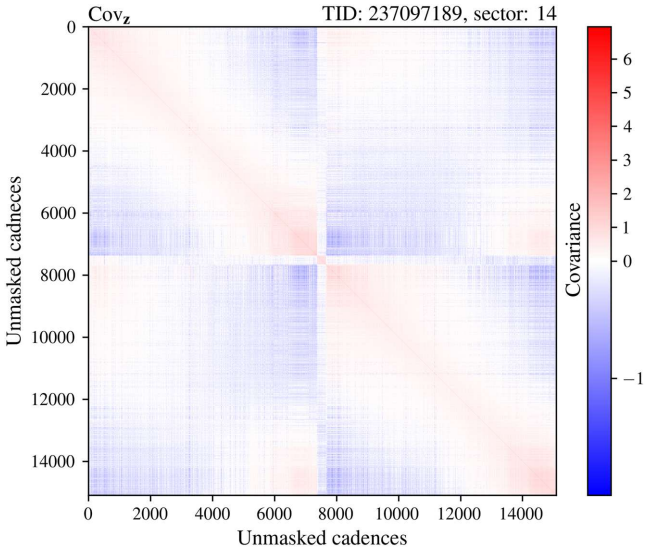



**Figure 8.** A sample joint-covariance matrix is shown for a light curve. The noise exhibits a diagonal component of positive correlation due to short-term correlated noise described by $\boldsymbol{s}$, as well as systematic variability described by $\boldsymbol{l}$. Longer-term correlations are also seen with a 13.7 days blocked structure due to the dominant systematic, the 13.7 days downlink cycle. The dominant systematic is driven by the 13.7 days period orbit of TESS about the Earth. During the orbit the spacecraft experiences thermal changes and varying degrees of scattered light from the Earth and the Moon.

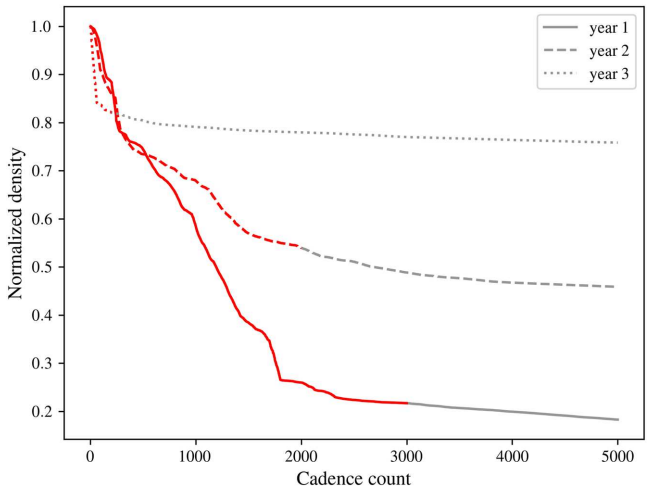



**Figure 9.** Sorted skyline histogram of the relative contribution of individual cadences to the detection statistics (as a normalized density) for all light curves. The $x$-axis markers indicate the number of individual cadences to the left of each marker, sorted in order of transit detection contribution (density). Separate histograms are plotted for the year 1, 2, and 3 data. The sort order of individual cadences by density differs by year. The red part of the curve indicates the knee cutoff used to mask cadences with anomalous density, and therefore likely unmodeled residual systematic errors. More cadences are filtered in the year 1 and 2 data compared to year 3.

epoch when finding the maximum detection statistic within this window. Transit events were not allowed to move relative to one another in a sector, however. This approach is also more robust to transit timing variations. In prior work (J. S. Taaki et al. 2020), we used the phase-correlation method to find the best-matched epoch $e$ of the template light curve. Here we do not utilize this method due to the phase-correlation method requiring multisector stitching and gap filling of data.

### 2.2.6. Postprocessing

In this application of our algorithm (J. S. Taaki et al. 2020) to TESS data we have improved robustness against FA detections. The detector form (Equation (5)) may produce spuriously high detection statistics (S. Seader et al. 2013; J. D. Twicken et al. 2016) in the presence of light-curve outliers, non-Gaussian noise, unmodeled residual systematics, or astrophysical mimics of exoplanet transits including EBs and stellar harmonics. For TESS, unmodeled residual systematics are significant across the data gaps at 13.7 days and 2–6 days, especially due to the data downlink and momentum dump cycle, respectively. In this section, we describe the postprocessing and filtering used to reduce FAs.

*Correlation statistic*. As a first step, we computed the correlation coefficient $r_{\boldsymbol{t}\boldsymbol{y}'} = \frac{\boldsymbol{t}^T \boldsymbol{y}_i'}{\| \boldsymbol{t} \| \| \boldsymbol{y}_i' \|}$ between the identified transit signal $\boldsymbol{t}$ and the associated processed light curve $\boldsymbol{y}_i'$. These light curves were processed using least-squares detrending against the SPOC CBVs (Section 2.2.3) but here also including the SPOC spike vectors (J. M. Jenkins et al. 2016). In addition, further masking was applied to these processed light curves $\boldsymbol{y}_i'$ beyond that described in Section 2.2.2. Specifically, cadences within 1 hr of momentum dump times were masked. In addition, a skyline histogram (Figure 9) was used to identify individual cadences contributing disproportionately to transit event detections in multiple light curves and therefore likely to contain unmodeled residual systematics.[12] As shown in red in this figure, approximately 1000–3000 cadences were masked for each year of data because of their anomalous contribution to detection statistics. Figure 10 shows the mean (in blue) of the detrended and masked light curves for all detections in the period range [1, 20] days in the sector 1 data overlaid with a detection density plot per cadence (gray); points filtered by the skyline histogram analysis are shown in red.

The correlation coefficient was computed within a window of tolerance about the expected location of a transit event, as

[12] github.com/christopherburke/TESS-ExoClass

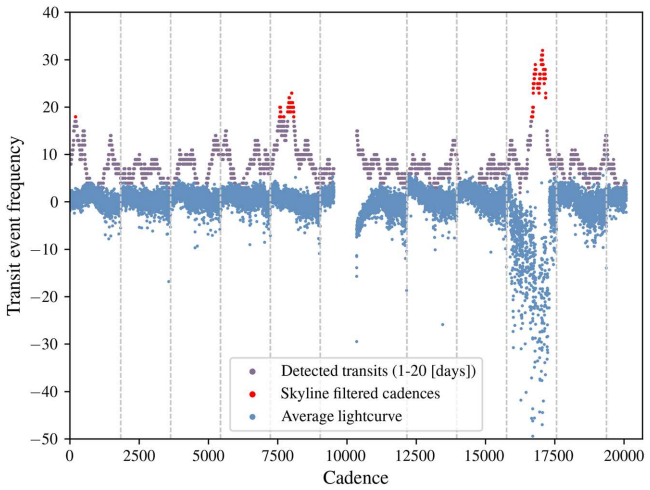



**Figure 10.** A skyline plot of the mean light curve (blue) here formed as the average over all detrended and masked detection light curves with a period in the range [1, 20] days in the sector 1 data. The vertical dashed gray lines indicate momentum dump events. The purple light curve is proportional to the frequency with which each individual cadence appears in any detected transit. Those cadences marked in red were filtered based on a skyline histogram analysis (Figure 9). Significant uncorrected systematics are visible in the cadence range 16,000–17,500, as evident in both the mean light curve (blue) and filtered cadences (red). However, overall for this period range the density of detections is not attributable to a single systematic source.

described in Section 2.2.5. All detections with a correlation metric $r_{ty'} < 0.1$ were filtered out as potential FAs. This removed approximately one-third of initial detections.

*Robustness statistic.* A chi-squared statistic, formed from constituent elements of the total detection statistic (S. Seader et al. 2013), is used in the TESS SPOC pipeline to filter FAs. We construct a similar, although simplified, robustness metric in this work, based on the principle that a true astrophysical transit should have consistent depth across all individual transit events.

We derive this robustness metric for our detector following the approach of S. Seader et al. (2013) but without their use of adaptive noise estimation or a multiscale wavelet decomposition. Each transit signal $\boldsymbol{t} = \sum_s \boldsymbol{t}_s$ can be decomposed as the sum of single-transit events $\boldsymbol{t}_s$ which are zero outside each individual transit domain.

The detection statistic is invariant to the choice of template transit depth (Appendix B). For a chosen transit template $\boldsymbol{t}$ of fixed depth and other transit parameters, we assume the light curve contains a true transit $\tilde{\boldsymbol{t}}$ described by a scaling $\alpha = \frac{\|\tilde{\boldsymbol{t}}\|}{\|\boldsymbol{t}\|}$ of the template transit $\boldsymbol{t}$ of the form $\hat{\boldsymbol{y}}_i = \alpha \sum_s \boldsymbol{t}_s + \boldsymbol{z}_i$.

From Equation (5) the single-transit detection statistic therefore has the following distribution for each hypothesis (S. Seader et al. 2013):

$$T(\hat{\mathbf{y}}_i; \mathbf{t}_s|_0) \sim \mathcal{N}(0, 1) \tag{8}$$

$$T(\hat{\boldsymbol{y}}_i; \boldsymbol{t}_s)|_1) \sim \mathcal{N}(\alpha \cdot m_s, 1), \tag{9}$$

where $m_s = \sqrt{\boldsymbol{t}_s^T \mathrm{Cov}_z^{-1} \boldsymbol{t}_s}$. Given the form of these distributions and allowing $\alpha \geqslant 0$ to describe both the null and alternate hypotheses, the negative log likelihood of the observed single-transit detection statistics for a candidate transit $\boldsymbol{t}$ can be expressed in the relation

$$-\ln\left[\mathcal{L}\left(\prod_s p(T_i(\hat{\boldsymbol{y}}_i; \boldsymbol{t}_s) = \bar{\alpha} \cdot m_s)\right)\right] \propto \frac{1}{N_s}\sum_s^{N_s}(T_i(\hat{\boldsymbol{y}}_i; \boldsymbol{t}_s) - \bar{\alpha} \cdot m_s)^2, \tag{10}$$

where $\bar{\alpha}$ is robustly estimated as the sample median of $\left\{\frac{T_i(\hat{\boldsymbol{y}}_i; \boldsymbol{t}_s)}{m_s}\right\}$, and $p$ denotes probability.

The offset statistic $T_i(\hat{\boldsymbol{y}}_i) - \bar{\alpha} \cdot m_s$ has a standard normal distribution, therefore the scaled sum of the variables in Equation (10) is theoretically $\chi^2(N_s)$, with an expected value of 1 (S. Seader et al. 2013). A lower value of the negative likelihood indicates that the light curve is well described by the hypothesis model for a transit (Equation (2)). However, the metric may exceed the theoretical distribution if the noise is non-Gaussian or if the search space discretization causes variable mismatch between the candidate and true transits. In particular, shorter-period transits will be subject to greater mismatch error.

We applied this statistic to long-period signals with fewer than five transits because some of these detected transits were found to contain a small number of cadences of significant outlier noise. In this case we rejected transit signals for which the sample variance of the detection statistic across all individual transits in the signal exceeded 1.5. Furthermore, over all candidate detections we rejected transits where $\hat{\alpha} < 0.0001$. In aggregate, these metrics removed ∼10% of targets with detections remaining after filtering by the correlation statistic.

*Eclipsing binary stars.* Astrophysical FAs can also be produced by EBs (A. Santerne et al. 2013). A catalog of candidate EBs was produced from the first 2 yr of TESS data by A. Prša et al. (2022).[13] We rejected candidate exoplanet detections if the target has been designated as a candidate EB, excluding 107 of our targets. As in the search by A. B. Justesen & S. Albrecht (2021) and A. Prša et al. (2022), we produced even–odd phase-folded light curves to identify EBs from the ratio of odd and even transit depths. The flagged EB candidates have a ratio of odd to even transit depths in excess of 10%. We use this as a coarse filtering metric here, and defer to future work the incorporation of transit depth uncertainty as used in J. D. Twicken et al. (2018).

### 2.3. Injection Tests

As in J. S. Taaki et al. (2020), injection tests were performed using the joint detector and a reference standard sequential detector to assess first-order algorithm performance. Injection tests are commonly used in this context (R. L. Gilliland et al. 2000; D. T. Weldrake et al. 2005; C. J. Burke et al. 2006; J. L. Christiansen et al. 2013, 2015, 2016; D. Foreman-Mackey et al. 2015).

For these tests a total $N = 1000$ light curves were randomly selected from the target selection of year 1 data from the CVZ; these comprise a significant fraction of the available data. A transit signal for a single exoplanet was injected into each selected light curve. The transit signals were simulated using the `transit` software library,[14] which includes limb

[13] tessebs.villanova.edu
[14] http://dfm.io/transit

**Table 2**
Injected Signal Parameter Distribution

| Transit Parameter | Distribution |
|---|---|
| Period $P$ (days) | $U(1, 100.)$ |
| Radius ratio of planet to host star (%) | $U(0.01, 0.2)$ |
| Transit epoch $e$ (days) | $U(0, P)$ |
| Impact parameter (stellar radii) | $U(0, 1)$ |
| Argument of periapse $\omega$ (rad) | $U(-\pi, \pi)$ |
| Limb-darkening parameters: $q_1$, $q_2$ | $U(0, 1)$ |

**Note.** The distribution of injected signal parameters. A uniform probability density function over the domain $\{x_1, x_2\}$ is denoted as $U(x_1, x_2)$. See D. M. Kipping (2013a) for a definition of limb-darkening parameters.

darkening (K. Mandel & E. Agol 2002; D. M. Kipping 2013a) and Keplerian orbital dynamics. The transit parameters of the injected signals were drawn from a representative exoplanet population (Table 2) informed by prior population characterization (D. M. Kipping 2013a, 2013b; D. Foreman-Mackey et al. 2015).

The reference detector was implemented as described in J. S. Taaki et al. (2020). This detector has standard detrending and detection phases. The detrending assumes no transit signal is present, $\boldsymbol{y}_i = \boldsymbol{s}_i + \boldsymbol{V}\boldsymbol{c}_i$ (hypothesis $H_0$; Equation (2)), and $\boldsymbol{V}$ is spanned by basis vectors $\boldsymbol{v}_k$ (Section 2.1), which are here the TESS single-scale CBVs (as in Section 2.2.3). The estimated basis vector coefficients $\hat{\boldsymbol{c}}_{0,i}^{\mathrm{MAP/MMSE}}$ were obtained using a MAP estimator (here equivalently a minimum mean square error estimator) and using systematic and statistical priors obtained from the injected light curves utilizing the same procedure described in Sections 2.2.3 and 2.2.4.

The estimator can be derived either by the expectation with respect to the posterior distribution $\mathbb{E}_{H_0}(\boldsymbol{c}_i|\boldsymbol{y})$ or, as shown by J. C. Smith et al. (2012), by maximizing the log of the posterior $\arg\max_{\boldsymbol{c}_i} \ln p_{H_0}(\boldsymbol{c}_i|\boldsymbol{y}_i)$:

$$\hat{\boldsymbol{c}}_{0,i}^{\mathrm{MAP/MMSE}} = (\boldsymbol{V}^{\boldsymbol{T}}\mathrm{Cov}_{s,i}^{-1}\boldsymbol{V} + \mathrm{Cov}_{\boldsymbol{c}}^{-1})^{-1}(\boldsymbol{V}^{\boldsymbol{T}}\mathrm{Cov}_{s,i}^{-1}\boldsymbol{y}_i + \mathrm{Cov}_{\boldsymbol{c}}^{-1}\mu_{\boldsymbol{c}_i}). \quad (11)$$

The light curves were then detrended as $\boldsymbol{y}_i^{co} = \boldsymbol{y}_i - \boldsymbol{V}\hat{}_{i,0}^{\mathrm{MAP/MMSE}}$. The sequential detection step was performed using the detection statistic (Equation (5)) decomposed over sector (Appendix B), with $\mathrm{Cov}_{z,i}$ replaced by $\mathrm{Cov}_{s,i}$ as in J. S. Taaki et al. (2020).

No postprocessing or filtering was performed for the injection tests because their purpose was solely to characterize the comparative performance of the joint and sequential detection approaches. The detection statistics are implemented as in Section 2.2.5, allowing candidate transits over a sector to shift by $\pm 2$ hr. A candidate detection was recorded for an injected light curve if the detection statistic exceeded the threshold $T(\hat{\boldsymbol{y}}_i) > 10$ for a candidate transit signal $\boldsymbol{t}$ and considering only the detected candidate transit signal $\boldsymbol{t}$ with the maximum test statistic. The number of light curves with detections is denoted $N_d$. The true injected transit signal $\boldsymbol{t}_{\mathrm{inj}}$ in a light curve was considered to be correctly recovered the candidate transit signal $\boldsymbol{t}$ if either (i) the orbital periods of the detected $P(\boldsymbol{t})$ and injected $P(\boldsymbol{t}_{\mathrm{inj}})$ transit signals matched within $|P(\boldsymbol{t}) - P(\boldsymbol{t}_{\mathrm{inj}})| < 2$ hr, or (ii) the correlation metric between the injected and detected transit signals exceeded 0.75. As described in Section 2.2.5, the transit template is allowed to shift $\pm 2$ hr, and this was similarly applied when computing the correlation metric. Furthermore, a box approximation was used for both injected and detected transits. Criterion (ii) was added because for long-period (>15 days) injected transit signals a significant proportion of the transits may occur on data gaps. Therefore, a transit is considered to be detected if the majority of transit events are recovered; however, it is not required to match in detected period per criterion (i). Since the quasi-FA rate is intended to provide a measure of erroneous detections from systematics, we do not count detections which are correlated with the injected transit as FAs. A correlation threshold of 0.75 was chosen as this implies that more than half of detected and injected individual transits match.

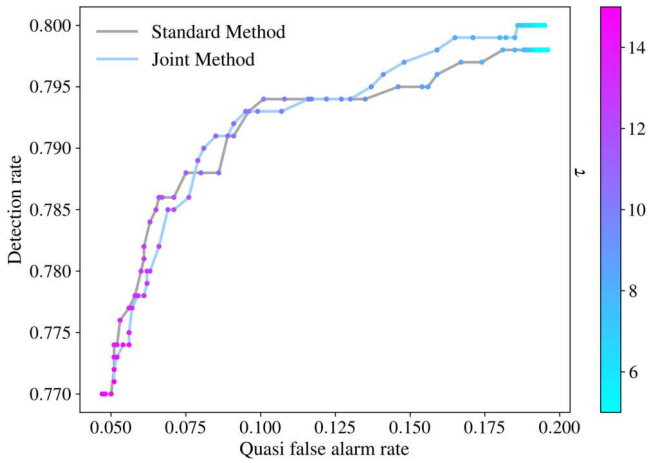



**Figure 11.** Injection test detection rate vs. quasi-FA rate over detection threshold $\tau$ (right color wedge) for the joint and standard reference detection algorithms. The two methods are comparable in performance, showing only a marginal improvement in detection power for the joint method.

The number of injected light curves with correct detections meeting these criteria is denoted $N_c$. The detection rate is $R_D = \frac{N_c}{N}$. The quasi-FA rate $R_{\mathrm{QFA}} = \frac{N_d - N_c}{N}$ measures the rate of incorrect detections. We note that a FA may in fact be a true astrophysical transit present in the TESS light curve before the injected transit signal was added; this particularly affects detection rate estimates for weaker injection transits. However, this is mitigated by the low expected density of preexisting transits in the set of injected light curves and the higher strength of injected transits in general relative to true preexisting transits. Therefore, we believe that the quasi-FA rate $R_{\mathrm{QFA}}$ reasonably identifies noise properties or features that pose challenges to correct detection.

## 3. Results

### 3.1. Injection Test Performance

The injection tests used to assess algorithm performance are described in Section 2.3. In these tests the standard sequential detector used as a reference achieved a detection rate $R_D = 79.8\%$ and a quasi-FA rate $R_{\mathrm{QFA}} = 19.1\%$. The joint detector (J. S. Taaki et al. 2020) as implemented here achieved a marginal increase in detection rate of 0.2%, although this is not statistically significant given the sample size. The detection rate $R_D$ is plotted in Figure 11 against quasi-FA rate $R_{\mathrm{QFA}}$ as a function of detection threshold $\tau$ for both the joint and the standard reference algorithms. The detection rate of the joint detector at $\tau = 10$ is shown in Figure 12, tabulated over orbital

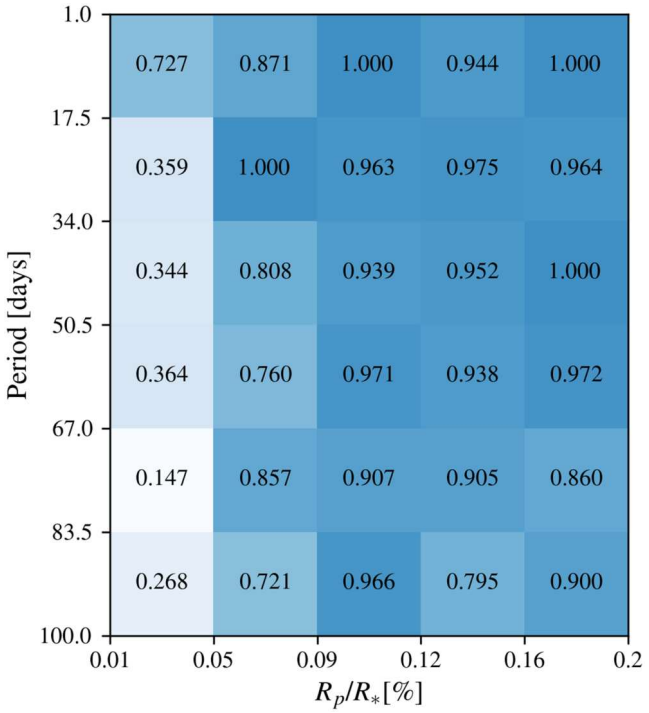



**Figure 12.** The detection rate of the joint detector at a threshold of $\tau = 10$ for transit injection tests with 1000 light curves from year 1, tabulated over orbital period and planet-to-star ratio $R_p/R_*$. The distribution of simulated transit signal parameters can be found in Table 2. The mean detection rate and quasi-FA rate are $R_D = 80.0\%$ and $R_{\rm QFA} = 19.1\%$, respectively.

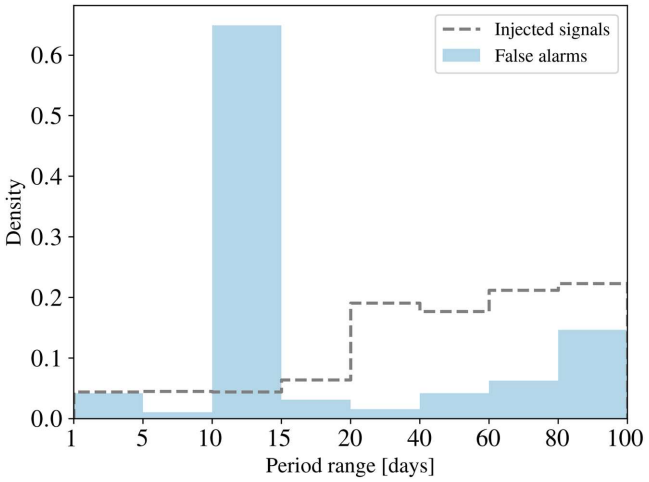



**Figure 13.** For the injection tests, the distribution of quasi-FA detections (188/1000, blue) and injected transits (gray dashed) as a function of their detected or injected periods respectively are plotted as a normalized density. Half of the FA detections have a period between 10 and 15 days, likely due to the 13.7 days downlink cycle. The density of FAs further increases with period, likely due to outlier scatter. We note that the bin size on the $x$-axis is not uniform.

period and the planet-to-star radial ratio $R_p/R_*$ of the injected signals.

For the injection tests, the distribution of quasi-FAs and correct detections over detected period is shown in Figure 13. Over 60% of the quasi-FAs occur due to unmodeled systematics or scatter spanning the 13.7 days downlink cycle gap and also possibly associated with the 2–6 days momentum dump cycle. In addition, the quasi-FA rate increases with period; this is likely due to outlier events remaining in the data.

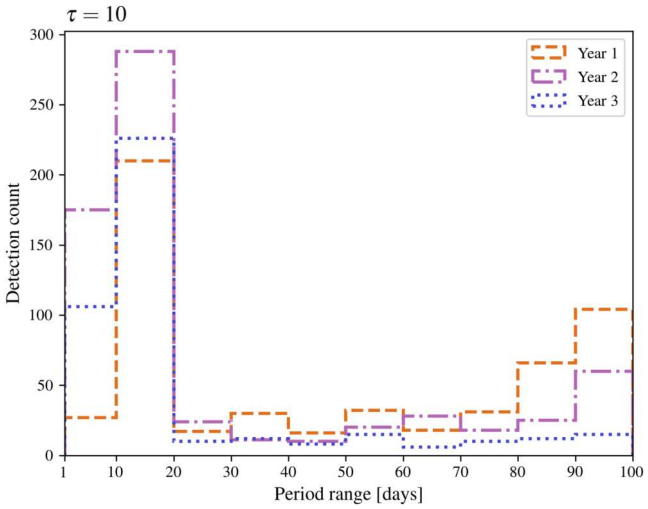



**Figure 14.** The distribution of detections in the TESS data search (Section 3.2) each year over period. As in the injection tests (Figure 13) there is an increase in suspected FA detections for periods $P < 20$ days, likely due to residual systematic scatter.

### 3.2. TESS Search

An important goal of this work is to perform an independent search of the first 3 yr of the TESS 2 minutes SAP light curves from the CVZ using the joint detector (J. S. Taaki et al. 2020) as implemented here. No light curves with known TESS TOIs produced by the SPOC/QLP pipelines were excluded from our search in advance. We compare our recovery rate of known TOIs as a coarse form of validation. We note that the period range searched by the SPOC/QLP pipeline differs from ours, and by definition any candidate transits lying outside of our search range are excluded from the listed TOIs here. Postprocessing and filtering (Section 2.2.6) were applied to our candidate detections, which were obtained at a detection threshold $\tau = 10$. The detection threshold was set based on the injection tests balancing the FA rate against sensitivity to weak transit signals. As noted in Section 2.2.6, postprocessing includes the removal of known or suspected EBs. A total of 78 candidate detections were removed as known EBs cataloged in A. Prša et al. (2022), including three identified by SPOC and one listed in T. Mitnyan et al. (2020). We note that a further 21 known EBs are cataloged for light curves in which we did not detect any transits. The majority of missed EBs have very short orbital periods <2 days, with a large morphology parameter $c > 0.5$ (G. Matijevič et al. 2012). These high-contact systems have phase-folded light curves with wide eclipses, unlike transiting exoplanets.

After postprocessing and filtering, approximately 28% of light curves contained a detection, with 551/1851 in year 1, 659/2046 in year 2, and 420/1922 in year 3. The distribution of these detections over period is shown in Figure 14 for each year separately. This figure confirms the quasi-FA distribution with period found in the injection tests (Figure 13). The postprocessing and filtering is somewhat robust to outlier signals, however these processes are not robust to unmodeled quasi-periodic systematics that may mimic transit signals. This is particularly true for those associated with the 13.7 days downlink and 2–6 days momentum dump cycles, as suggested in these figures. To explore this further, the detection count in the TESS data search is tabulated in Figure 15 over detected period $P$ (for $P < 19$ days) and detection-statistic value $10 \leqslant T(\hat{\boldsymbol{y}}) \leqslant 20$ for all 3 yr of TESS data combined. This

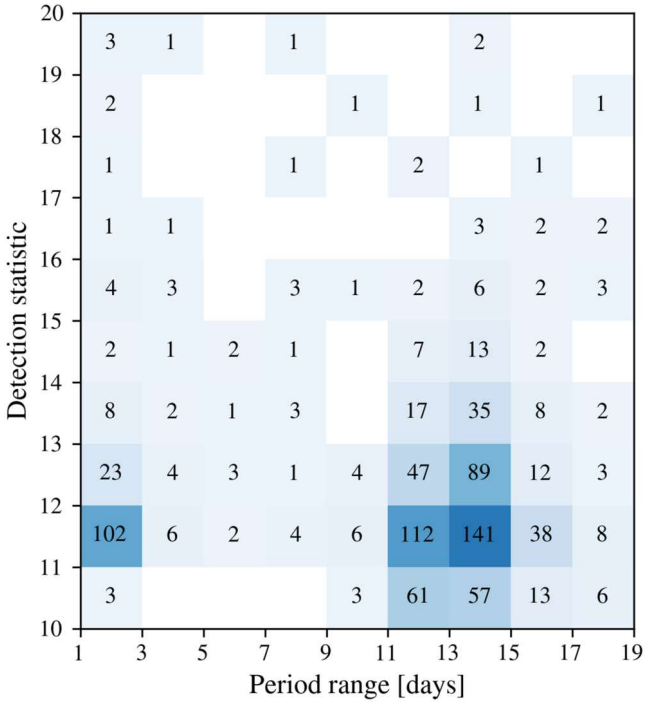



**Figure 15.** Detection counts for the TESS data search tabulated over detected period ($P \leqslant 19$ days and detection test statistic $10 \leqslant T(\hat{\boldsymbol{y}}) \leqslant 20$)) for all 3 yr combined. The overdensity of detections between 1 and 3 days and 11–15 days are very likely FAs due to residual systematic scatter. Both period ranges map directly to mission instrumental cycles, as described in Section 3.1.

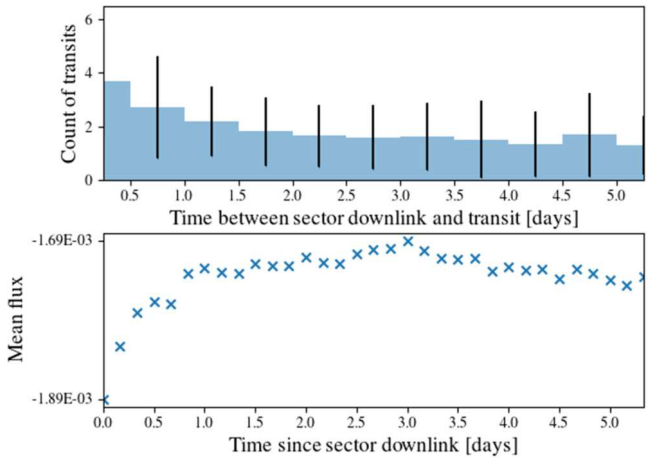



**Figure 16.** In the upper panel, the count of transit events binned in 0.5 days time-offset intervals between the sector downlink event and each transit event, for year 2 of detected transits and including all sectors. The error bar represents the standard deviation of the transit count over the set of light curves. In the lower panel, the median light-curve flux density in the same time-offset bins is shown at a bin width of 4 hr. Increased negative scatter occurs close to the sector downlink times and is correlated with an increased transit event count, suggesting the many detections with periods between 10 and 15 days are likely FAs.

provides further support for this hypothesis. Further supporting evidence is provided by Figure 16, which shows a strong correlation between detected transits and sector downlink edges and also shows that these are regions of higher systematic light-curve scatter.

Approximately 63% of known TOIs were recovered in these detections as described below in Section 3.2.1. Informed by the quasi-FA distribution over period discussed above, the majority of exoplanet candidates with periods between 10 and 15 days were excised during this manual vetting. Similarly, there was careful vetting of periods in the 2–6 days interval due to the momentum dump cycle and the overlapping 1–5 days interval quasi-periodic stellar variability known in the 1–5 days interval from Kepler threshold-crossing events (TCEs) processing (S. E. Thompson et al. 2018). Additional manual vetting of phase-folded light curves was performed. However, MAST SPOC data validation (DV) reports indicated that the remaining candidate detections were most likely EBs not previously identified as TCEs.

#### *3.2.1. TOI Recovery*

TESS project TOIs are produced by filtering TCEs identified by the QLP/SPOC pipeline. Automated and manual vetting is applied in this process to reduce FAs and FPs (N. M. Guerrero et al. 2021). We downloaded the project TOI list from the ExoFOP database.[15] As noted under the TOI release notes,[16] there may be inconsistencies between the fitted transit parameters across the TOI lists published by the TESS Science Office, MAST, or ExoFOP. We note that only a small fraction (∼12%) of TOIs are, as yet, confirmed exoplanets (E. J. Melton et al. 2024), as reasonably expected.

Given this heterogeneity, we used weaker criteria for TOI recovery than those used in the injection tests (Section 2.3). For each light curve we consider a detected transit signal $\boldsymbol{t}$ to recover a known TOI $\boldsymbol{t}_{\rm TOI}$ associated with the target if (i) the detection statistic for $\boldsymbol{t}$ exceeds the detection threshold $\tau = 10$, (ii) the detection statistic for $\boldsymbol{t}$ is the maximum for the light curve, and (iii) the period of the detection $P(\boldsymbol{t})$ matches the period of the TOI $P(\boldsymbol{t}_{\rm TOI})$ within a tolerance $|mP(\boldsymbol{t}) - P(\boldsymbol{t}_{\rm TOI})| < 2$ hr $(m \in \{2, 3, \frac{1}{2}, \frac{1}{3}\})$; the integer multiplicity in the allowed period match addresses a small number of cases (<5) in which the detected or TOI periods may have a low-order integer ambiguity due to data gaps, as noted for the QLP (N. M. Guerrero et al. 2021). These three criteria define the TOI match or recovery rate in what follows. If the detection statistic for $\boldsymbol{t}$ is above the threshold but not a maximum (i.e., only (i) and (iii) are satisfied), then this defines the TOI recall rate as used here. We note that 12 light curves contained more than one TOI; we count these as a single TOI in the total TOI count, a singular recall if any TOIs are among the detections, and a single match if any of the TOIs are recovered according to the detection criteria.

Over the 3 yr of data, we found a 63% TOI recovery rate and a 73% TOI recall rate. Table 3 shows for each year the number of light curves with known TOIs, the recall and match rates for these TOIs, and a separate match rate excluding TOIs designated by TESS as potential FPs or FAs. The TOI count and matched TOI detections are shown over period in Figure 17. This figure shows a slight improvement in TOI recovery for periods between 10 and 100 days.

## 4. Discussion

A goal of the current work is to evaluate the feasibility of applying our joint detector, previously developed and evaluated with Kepler data (J. S. Taaki et al. 2020), to data

[15] https://exofop.ipac.caltech.edu/tess/
[16] https://tess.mit.edu/toi-releases/

**Table 3**
TOI Detections at $\tau = 10$

| | Total TOIs | TOI Recalls | Recall Rate | TOI Matches | Match Rate | Match Rate ($\overline{\text{FP}}$, $\overline{\text{FA}}$) |
|---|---|---|---|---|---|---|
| Year 1 | 30 | 20 | 67% | 17 | 57% | 55% |
| Year 2 | 21 | 18 | 86% | 15 | 71% | 68% |
| Year 3 | 47 | 34 | 72% | 30 | 64% | 64% |
| Overall | 98 | 69 | 73% | 62 | 63% | 62% |

**Note.** From left to right, the columns are year number, the total number of light curves with TOIs present, the number of TOIs recalled, the associated recall rate, the number of TOI matches or recoveries, the associated match rate, and the match rate with any TOIs that are designated by the TESS project as false positives (FPs) or false alarms (FAs) excluded. The final row gives the combined statistics across years 1 to 3.

from the TESS mission, another wide-field exoplanet transit survey with different instrumental systematics and observing cadence. We find the modifications and enhancements required to be algorithmically feasible (Section 2.2) and computationally tractable (Appendix B.1), albeit remaining in the HPC domain (J. S. Taaki et al. 2020).

An additional goal of the paper is to assess the statistical performance of the joint detector in the new application to TESS data. In transit injection tests (Section 3.1) we find the joint detector to be comparable to or slightly exceed (Figure 11) the statistical performance of a standard sequential detrending and detection algorithm, used here as a reference. At a detection threshold $\tau = 10$, the joint detector achieved a detection rate $R_D = 80.0\%$ at a quasi-FA rate $R_{\text{QFA}} = 19.1\%$ for $N = 1000$ injected transit signals. For a fixed $R_{\text{QFA}}$, this is a $\triangle R_D = 0.2\%$ improvement over the reference sequential detector; however, this is not statistically significant. We note an important caveat that the standard sequential detector (Section 2.3) is implemented as a modification of our pipeline and is only representative of this class of sequential detector; we make no claim of optimality relative to other implementations.

In prior work with Kepler data (J. S. Taaki et al. 2020) the joint detector was found to have an advantage in detecting short-period, low-transit-depth exoplanets relative to a standard sequential detector, likely by mitigating overfitting. In the current work with TESS data, no such advantage is evident. In this work, we have used the well-constrained TESS project CBVs for all detectors, thereby possibly minimizing overfitting by the sequential detector. The standard deviation of each cadence under the systematics noise prior is of comparable magnitude to the stochastic noise, which includes stellar variability (Figure 7). In addition, the single most likely sources of residual systematic error, and therefore FAs, are the 13.7 days instrument downlink cycle and the 2–6 days momentum dump cycle, as discussed in further detail below. This effect is sufficiently prominent that it may have overshadowed more subtle detection differences evident in the prior Kepler work (J. S. Taaki et al. 2020), limiting our conclusions about joint detection here. Our injection tests provide insight into the algorithmic choices for these specific data: They do not necessarily generalize to other data sets or address the global optimality of joint modeling. This question is broad (D. Foreman-Mackey et al. 2015; G. Kovács et al. 2016; L. J. Garcia et al. 2024) but remains unsettled in the literature. Notably, L. J. Garcia et al. (2024) identify a benefit of their approach compared to other common transit search implementations for rapidly rotating M dwarfs.

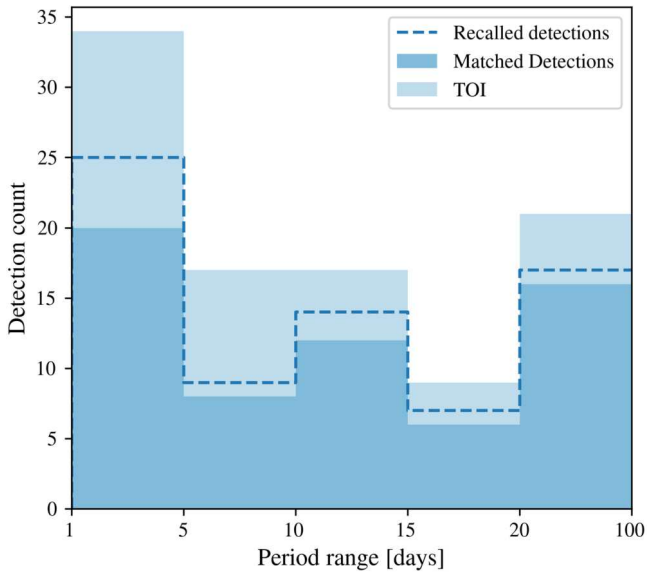



**Figure 17.** The distribution of known TOIs (light blue), known TOIs recovered as the top detection by the joint detector (darker blue), and known TOIs recovered as any detection above threshold by the joint detector (dashed line) over period. The TOI recovery rate increases slightly in the period range 10–100 days. We note that the bin size on the $x$-axis is not uniform.

We find the statistical error of the joint detector applied to TESS data to be strongly dependent on exoplanet orbital period $P$. Specifically, the transit injection tests show a significantly enhanced quasi-FA rate for the detected orbital period range $P \in [10, 15]$ days and a secondary enhancement at $P < 5$ days (Figure 13). The full search of the input TESS data (Section 3.2) similarly shows (Figures 14 and 15) an anomalous excess of detections in the same period range, with a secondary detection excess in the range $P < 5$ days, both likely quasi-FAs (Figure 13). There is evidence that unmodeled systematics near the 13.7 days instrument downlink data gap are responsible for the excess detections for $P \in [10, 15]$ days (M. Kunimoto et al. 2023), and also as examined in the analysis in Figure 16. Furthermore, the spurious detections for $P \in [2, 5]$ days likely occur due to unmodeled systematics associated with the momentum dump cycle, also noted as a source of FPs in the TESS TOI search (N. M. Guerrero et al. 2021). The robust CBVs used do not include edge flux ramps prior to downlink, as these are typically corrected via spline regression in an independent part of the SPOC pipeline (J. M. Jenkins 2017). Edge correction may help to mitigate the density of these spurious detections. However, the SPOC pipeline also produced increased FAs between $P \in [10, 15]$ days (M. M. Fausnaugh et al. 2018) due to downlink scatter. The sensitivity of the statistical error to exoplanet period is not specific to the joint detector. It occurs with other transit detectors, such as the reference sequential detector used in this work, and is foundationally related to the TESS observing and operations schedule structure.

The injection tests (Figure 13) also show an increase in the quasi-FA rate for longer periods $P > 60$ days, as also shown in the full TESS data search (Figure 14). We believe that these FA detections occur due to nonperiodic instances of outlier scatter, particularly for the year 1 data (on which the transit

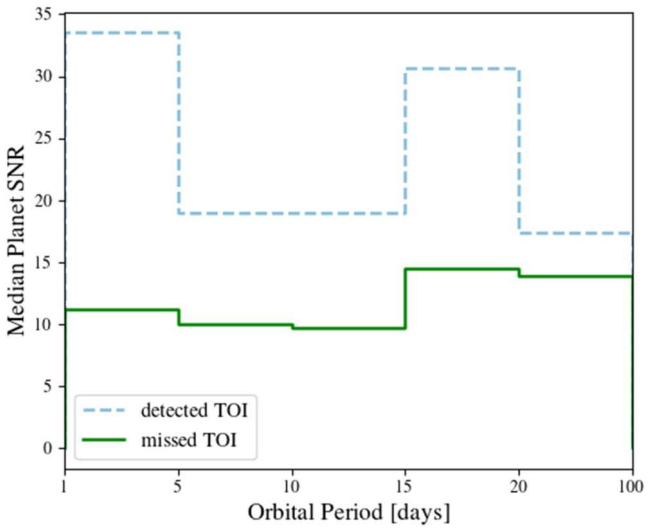



**Figure 18.** The distribution of the median planet signal-to-noise ratio (SNR) of missed TOIs (green line) and detected TOIs (dashed blue line) over period, as obtained using the joint detector. The planet SNRs are those provided by ExoFOP. Although the detection thresholds cannot be compared directly, our detection threshold ($\tau = 10$) exceeds the MES detection threshold of 7.1, and this may reasonably contribute to differences in candidate detections. In general, the missed TOIs have a lower planetary SNR of 11.3, compared to a value of 23.0 for the detected TOIs.

injection tests were performed). This likely reflects an inadequacy in our stochastic noise estimator, specifically a failure to model local noise variability (Section 2.2.6) and the assumption of stationarity over a sector. This suggests that future work to incorporate local downlink and outlier scatter in a stochastic noise prior using window estimation akin to J. M. Jenkins et al. (2010) may be a productive approach in future work. Furthermore, when combining multisector transit detection statistics, the $\pm 2$ hr epoch shift tolerance may contribute instances of FAs with longer orbital periods.

The input light curves in the full TESS search (Section 3.2) contain 98 unique targets with known TOIs (Table 3), provided by various detection pipelines including SPOC (J. Twicken et al. 2019) and the QLP (N. M. Guerrero et al. 2021). Under the definitions and criteria in Section 3.2.1, a recall rate of 73% and a recovery rate of 63% are broadly comparable to the TOI recovery rate of other detectors for which this metric has been evaluated (J. Chakraborty et al. 2020; N. L. Eisner et al. 2020). A priori, lossless TOI recovery is not expected given differences in algorithm approach, implementation, and the sources of FAs identified by the diagnostic tests described in Section 3.2. In particular, although the detection thresholds are not directly comparable quantities, our detection threshold $\tau = 10$ is higher than the project multiple event statistic (MES) detection threshold of 7.1 (M. Fausnaugh et al. 2021; N. M. Guerrero et al. 2021). The distribution of TOI recovery rate over period $P$ (Figure 17) shows higher TOI recovery and recall rates for period $P > 10$ days and also greater scatter in these rates for $P < 10$ days. In contrast, the injection tests show a monotonic decrease in recovery rate with increasing period (Figure 12). This discrepancy is likely explained by the sample signal statistics of the TOIs. As shown in Figure 18, the TOIs with orbital period $P < 10$ days missed by the joint detector generally have a lower median planet signal-to-noise ratio reported on ExoFOP compared to both the detected TOIs and the missed TOIs with longer orbital periods. The missed TOIs have an overall median planetary SNR of 11.3 compared to the detected TOI value of 23.0.

A final goal of the current work is to conduct an independent search of the TESS 2 minute SAP light curves for the CVZ using the joint detector. Complementary exoplanet transit searches of these data are important to find possible missed exoplanet detections. After manual vetting to eliminate EBs, stellar harmonics, or noise features mimicking transits, and cross-referencing against the MAST SPOC DV reports for TCEs, we find none of the detected signals can be unambiguously classified as exoplanets. We note in closing that the results presented here depend on specific tunable parameter or implementation choices such as transit template search space discretization and chosen noise priors, among others. This work was not designed to fully optimize these choices. However, as noted in this section, existing work including the injection tests has identified several aspects of the pipeline that could be optimized to improve performance. As noted above, a specific example includes improved systematic and stochastic noise priors to better model downlink scatter and local noise variance.

## 5. Conclusions

In this work, we applied a prior joint detection method for exoplanet transits (J. S. Taaki et al. 2020), which uses Bayesian priors for systematic and stochastic noise and was previously applied to Kepler data, to the first 3 yr of TESS 2 minutes SAP data from the CVZ. An exoplanet search was conducted over periods between 1 and 100 days. We modified the detector implementation to address unique aspects of TESS data and to provide an efficient HPC implementation. The detector pipeline is robust to spurious harmonics occurring due to stellar variability, and gap filling or stitching is not explicitly required. We utilized several refined postprocessing methods to suppress FAs due to outlier noise and residual quasi-periodic systematics.

Our resulting pipeline performance was first evaluated using transit injection tests. The joint detector was evaluated against a sequential detrending and detection algorithm used as a reference standard implementation. We also evaluated the recovery of known TOIs present in the light curves by performing a full search on the TESS input data. The principal conclusions of this work are as follows:

1. The detector was successfully modified and applied to a new wide-field exoplanet transit mission (TESS). In optimized form it is computationally tractable but remains in the HPC domain.
2. The injection tests show a high detection rate $R_D = 80.0\%$ with a quasi-FA rate $R_{\rm QFA} = 19.1\%$ at a detection threshold $\tau = 10$. A marginal improvement in detection performance $\triangle R_D = 0.2\%$ is achieved relative to the reference sequential detector. We posit that the similar performance is primarily due to a systematic noise prior that is fairly well constrained relative to stochastic variability per light curve.
3. An excess of quasi-FAs is found in period ranges $P \in [10, 15]$ days and $P < 5$ days, most likely due to unmodeled systematics related to the 13.7 days downlink cycle and the 2–6 days momentum dump cycle intrinsic to the TESS operational schedule.

4. In our full-scale search of 3 yr of TESS 2 minute light curves from the CVZ, we recover 63% of known TOIs with a matching orbital period and achieve a 73% recall rate for TOIs when compared to all detected transits.

## Acknowledgments

We thank the anonymous referee for their insightful comments on this topic.

This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.

Some of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute (MAST Team 2021). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5–26555. Support to MAST for these data is provided by the NASA Office of Space Science via grant NAG5–7584 and by other grants and contracts. This paper includes data collected with the TESS mission. Funding for the TESS mission is provided by the NASA Explorer Program. STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5–26555.

This research has made use of the Exoplanet Follow-up Observation Program (ExoFOP; NExScI 2022) website, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

*Software:* This work used the following external packages: Astropy (Astropy Collaboration et al. 2013, 2018, 2022), Scikit-learn (F. Pedregosa et al. 2011), Matplotlib (J. D. Hunter 2007), NumPy (C. R. Harris et al. 2020), SciPy (P. Virtanen et al. 2020), and transit v0.3.0 (https://github.com/dfm/transit).

## Appendix A
## List of Symbols

A list of important symbols and mathematical notation used in the paper is provided in Table 4.

**Table 4**
List of Symbols

| | |
|---|---|
| $\hat{\boldsymbol{y}}$ | Light curve |
| $\boldsymbol{z} \sim \mathcal{N}(0, \mathrm{Cov}_{\boldsymbol{z}})$ | Joint noise model |
| $i \in I$ | Index set of unique targets |
| $q \in Q$ | Set of TESS observational sectors |
| $N$ | Length of a light curve |
| $N_q$ | Length of a single sector of a light curve |
| $d \in D$ | Set of discrete transit durations |
| $P$ | Transit orbital period |
| $e$ | Transit epoch |
| $\alpha$ | Relative transit depth |
| $\delta P$ | Discrete sampling increment in period |
| $\delta e$ | Discrete sampling increment in epoch |
| $\boldsymbol{t} \in \boldsymbol{T}$ | Transit template search space |
| $\boldsymbol{T}_d$ | Subspace of $\boldsymbol{T}$ for a transit duration $d$ |
| $\boldsymbol{T}_{d(q)}$ | Subspace of $\boldsymbol{T}$ for a transit duration $d$ and a single sector of light-curve data |
| $\mu_{\boldsymbol{t}}$ | Average number of transit events per $\boldsymbol{t} \in \boldsymbol{T}$ |

## Appendix B
## Sector-to-sector Detector

Analogous to the discussion in J. S. Taaki et al. (2020, Appendix B), we assume independent noise in each of the 13 sectors in each year-long light curve. Dropping the light-curve index $i$, systematics are estimated sector to sector, and a sector-long model for statistical noise $\boldsymbol{s}$ is also appropriate relative to transit timescales. For each sector $q \in Q$ the joint noise covariance is therefore denoted $\mathrm{Cov}_{z(q)} = \mathrm{Cov}_{s(q)} + \boldsymbol{V}_q \mathrm{Cov}_{c(q)} \boldsymbol{V}_q^T$.

Scaling the candidate transit signal $\boldsymbol{t}$ by $\alpha \in \mathbb{R}^+$ will not affect the matched filter statistic (S. Kay 1998). For a signal $\boldsymbol{t}_\alpha = \alpha \boldsymbol{t}$:

$$\frac{\hat{\boldsymbol{y}}^T \mathrm{Cov}_{\boldsymbol{z}}^{-1} \boldsymbol{t}_\alpha}{\sqrt{\boldsymbol{t}_\alpha^T \mathrm{Cov}_{\boldsymbol{z}}^{-1} \boldsymbol{t}_\alpha}} = \frac{\hat{\boldsymbol{y}}^T \mathrm{Cov}_{\boldsymbol{z}}^{-1} \alpha \boldsymbol{t}}{\sqrt{\alpha^2 \boldsymbol{t}^T \mathrm{Cov}_{\boldsymbol{z}}^{-1} \boldsymbol{t}}} = \frac{\hat{\boldsymbol{y}}^T \mathrm{Cov}_{\boldsymbol{z}}^{-1} \boldsymbol{t}}{\sqrt{\boldsymbol{t}^T \mathrm{Cov}_{\boldsymbol{z}}^{-1} \boldsymbol{t}}}, \tag{B1}$$

$$T(\boldsymbol{y}) = \frac{1}{\sqrt{Q}} \sum_{q \in Q} \frac{\hat{\boldsymbol{y}}_q^T \mathrm{Cov}_{z(q)}^{-1} \boldsymbol{t}_q}{\sqrt{\boldsymbol{t}_q^T \mathrm{Cov}_{z(q)}^{-1} \boldsymbol{t}_q}} \underset{H_0}{\overset{H_1}{\gtrless}} \tau. \tag{B2}$$

Therefore, a candidate transit depth is not needed to parameterize $\boldsymbol{t}$. However, to correctly compute a year-long detection test each sector light curve $\hat{\boldsymbol{y}}_q$ should be normalized so that transit strength is constant across all sectors. In our implementation, we chose to compute detection tests per sector and approximate a year-long detection test as the normalized sum of sector-wise matched filters described in Equation (B2). This average of detection tests is equivalent to a year-long detection test with the same noise model if the sector-to-sector noise $\mathrm{Cov}_{z(q)}$ is identically distributed between sectors. In informal tests, this detection statistic is found to be marginally lower, as the relative contribution of sectors with lower estimated noise levels is reduced.

### *B.1. Computational Methods and Complexity*

This section characterizes the computational complexity of the detection test kernel in the pipeline implementation (Figure 1). This work was completed with a 250,000 node-hour allocation on the Blue Waters petascale system, where 1 node-hour equals 32 core-hours. For each year of light-curve data, the detection tests required ∼50,000 node-hours of compute time and used ∼50 TB of intermediate storage.

#### *B.1.1. Efficient Computation of Multiple Transit Detection Tests*

This subsection concerns the computational optimization of detection-statistic calculations for transits of fixed duration but varying period and epoch. The specific parallel implementation of the detection pipeline is described in Appendix B.1.2.

The detection statistic (Equation (5)) includes the computation of a matrix-vector product $\mathrm{Cov}_{\boldsymbol{z}}^{-1} \boldsymbol{t}$ of order $O(N^2)$, where $N$ is the number of samples in the light curve. The inverse covariance calculations $\mathrm{Cov}_{\boldsymbol{z}}^{-1}$ are not a significant computational cost because they are independent of the transit signal $\boldsymbol{t}$ under study and are therefore calculated once per light curve. We optimize the computation of the term $\mathrm{Cov}_{\boldsymbol{z}}^{-1} \boldsymbol{t}$ over the transit search space $\boldsymbol{t} \in \boldsymbol{T}$ (Table 1) by retaining intermediate results exploiting similarities for transit signals of fixed duration $d$; a similar approach is described by J. M. Jenkins et al. (2010). A candidate transit signal $\boldsymbol{t}$ of duration $d$ may be decomposed into a number of disjoint single-transit events $\boldsymbol{t}_{d,s}$,

where $\mathrm{Cov}_z^{-1}\boldsymbol{t} = \sum_s \mathrm{Cov}_z^{-1}\boldsymbol{t}_{d,s}$. The matrix-vector product for a single-transit event $\mathrm{Cov}_z^{-1}\boldsymbol{t}_{d,s}$ involves the summation of $d$ matrix columns each of length $N$ which are contiguous in memory and can be accessed with an efficient cache hit rate; this specific computation consists of $d \cdot N$ additions.

For the transit template $\boldsymbol{t}_{d,s}$, we precompute $\mathrm{Cov}_z^{-1}\boldsymbol{t}_{d,s} \in \mathcal{R}^N$ with different positional shifts of the transit template over step sizes $\delta e$ in epoch. There are $\frac{N}{\delta e}$ possible positional shifts, and each calculation requires $Nd$ additions of the cached matrix $\mathrm{Cov}_z^{-1}$. Therefore, over all positional shifts a total of $\frac{N^2 d}{\delta e}$ additions are required. Over the full range of candidate durations $D$ these partial computations require $\sum_{d\in D}\frac{N^2 d}{\delta e}$ additions. These precomputed matrix-vector products are held in intermediate memory.

The transit search space $\boldsymbol{T}$ can be decomposed over subspaces of fixed transit duration $\boldsymbol{T}_d$ (Appendix D). For each transit signal $\boldsymbol{t} \in \boldsymbol{T}_d$, we compute $\mathrm{Cov}_z^{-1}\boldsymbol{t}$ by summing together the precomputed single-transit matrix-vector products. If the average number of transit events per candidate transit $\boldsymbol{t} \in \boldsymbol{T}_d$ is $\mu_t$ (Appendix D), then forming the sum of single-transit matrix-vector products to compute $\mathrm{Cov}_z^{-1}\boldsymbol{t}$ for all $\boldsymbol{t} \in \boldsymbol{T}$ requires $|\boldsymbol{T}|N\mu_t$ additions of the precomputed vectors given prior reduction over $d$.

The final terms to be formed in the detection statistic are the inner product calculations $\boldsymbol{y}^T\mathrm{Cov}_z^{-1}\boldsymbol{t}$ in the numerator and $\boldsymbol{t}^T\mathrm{Cov}_z^{-1}\boldsymbol{t}$ in the denominator. Given the precomputation of $\mathrm{Cov}_z^{-1}\boldsymbol{t} \in \mathcal{R}^N$ described above, forming these additional inner products requires $N$ multiplications and additions, respectively, for a total $|\boldsymbol{T}|N$ operations over $\boldsymbol{t} \in \boldsymbol{T}$.

Combining all complexity terms, the transit search of $\boldsymbol{T}$ for a single light curve $\boldsymbol{y}$ is of complexity $O\left(|\boldsymbol{T}|N(1+\mu_t) + \frac{\sum_{d\in D} dN^2}{\delta_e}\right)$. For the given transit search space, $|\boldsymbol{T}|\mu_t$ scales as $N^2$ and $\mu_t$ is constant (Appendix D). The leading-order complexity is therefore $O(|\boldsymbol{T}|N)$ or $O(N^3)$.

#### *B.1.2. High-performance Computing Pipeline*

SAP light-curve data are provided per TESS observational sector in the MAST archive. Each sector file is 1.5 MB in size and a year of light-curve data from the CVZ occupies ∼50 GB on hard disk. Detection statistics were computed per sector, as described in Section B.1.1, then combined across sectors (Equation (B2)) to compute detection tests for each light curve. We describe here the efficient implementation of these computations on the Blue Waters petascale system (C. L. Mendes et al. 2015). We note that detection-statistic computations are independent between light curves and between candidate transit durations.

The inverse covariance $\mathrm{Cov}_{z_i(q)}^{-1}$ of size $N_q \times N_q$ is first computed once per sector $q$ for each light curve, where $N_q$ is the number of points in the light curve for sector $q$. This covariance matrix is ∼1.7 GB in size and is stored on hard disk. For each light curve, the covariances for all 13 sectors occupy 22 GB on disk. All covariances for all light curves in the CVZ for a single year require ∼35 TB of disk storage.

A Blue Waters XE node consisted of 32 CPU cores and 32 GB of local memory. Our unit of parallelism is a task, and 20 tasks were deployed on each XE node. A task is the computation of a set of detection tests for $\boldsymbol{t} \in \boldsymbol{T}_{d(q)}$, where $\boldsymbol{T}_{d(q)}$ is the realization of the transit search templates in $\boldsymbol{T}_d$ for sector $q$. This transit search subspace is of size $\approx 10^7$ and holds transit templates of fixed duration $d \in D$ for a single light-curve sector $q$ of length $N_q$, over all period $P$ and epoch $e$ pairs in the sector for the parameter range defined in Table 1, and further described in Appendix D. Tasks for the same light curve $i$ and sector $q$ were assigned to the same node; there were $|D| = 10$ tasks per light curve. The common covariance matrix $\mathrm{Cov}_{z_i(q)}^{-1}$ was shared as a memory-mapped object between all tasks on the node. Each task is highly parallel, with no coupling or communication needed between tasks. In each task, single-transit event statistics of the form $\mathrm{Cov}_{z_i(q)}^{-1}\boldsymbol{t}_{d,s}$ were first computed serially. Only the $N_q \cdot d$ covariance matrix elements for each single-transit event $\boldsymbol{t}_{d,s}$ were read into local memory at a time. The memory required to hold all intermediate single-transit matrix-vector calculations $\mathrm{Cov}_{z_i(q)}^{-1}\boldsymbol{t}_{d,s}$ per task was $\frac{N_q^2}{\delta e}$ (Appendix B.1.1) 32 bit floating point values, of total size ∼0.4 GB. Within a single task, these single-transit calculations were combined to form detection tests in a serial manner for each period $P$ and epoch $e$ pair spanning $\boldsymbol{t} \in \boldsymbol{T}_{d(q)}$.

The output size of the detection tests of duration $d$ for a sector $q$ were $|\boldsymbol{T}_{d(q)}| \approx 10^7$ 32 bit floating point values, of total size 18 MB (therefore required $10^7$ writes per task).

The detection tests for a sector $q \in Q$ are intermediate as they are combined for all sectors to form composite detection statistics as described in Appendix B. The intermediate data output associated with a complete year of light-curve data, for all candidate durations $d \in D$, is of size $|\boldsymbol{T}_{d(q)}| \cdot |D| \cdot |Q| \cdot |I| \sim 15$ TB, where $|\boldsymbol{T}_{d(q)}|$ is the number of partial transits of duration $d$ and for a sector of data of length $N_q$, $|D| = 10$ is the number of searched candidate durations, $|Q| = 13$ is the total number of sectors spanning a year, and $|I| \sim 1500$ is the number of light curves per year of data. The detection test statistics computed from this intermediate data for all $\boldsymbol{t} \in \boldsymbol{T}$ for a single year-long light curve occupy 1 GB of total memory, and 5 TB for all light curves in the CVZ over 3 yr.

## Appendix C
## Loss in Detection Strength with a Box Transit Approximation

Approaches to bound the detection strength due to general template mismatch for a form of matched filter which depends on template amplitude is provided by H. L. Van Trees (2001, Section 4.2). Here we quantify the exact mismatch between a true transit function ($\boldsymbol{t}_d : \|\boldsymbol{t}\| = 1$) of transit duration $d$ and a box transit approximation ($\boldsymbol{b}_{d'} : \|\boldsymbol{b}_{d'}\| = 1$) of duration $d' \leqslant d$, where we assume all signals are centered on zero, for the form of matched filter used here which is invariant to template amplitude (shown in Appendix B):

$$\boldsymbol{b}_{d'}[n] = \begin{cases} \frac{1}{\sqrt{d'}} & \text{if } |n| \leqslant \frac{d'}{2}. \\ 0, & \text{otherwise} \end{cases} \tag{C1}$$

For mathematical convenience, we write the box transit normalized by duration as $\frac{1}{\sqrt{d'}}$; the detection statistics are invariant to this scaling. Consider the discretely sampled functions $\boldsymbol{t}_d$ and $\boldsymbol{b}_{d'}$ in vector form. The true transit signal $\boldsymbol{t}_d$ can be expanded as a linear combination along $\boldsymbol{b}_{d'}$ and an orthogonal vector $\boldsymbol{e}_d$, for which $\boldsymbol{b}_{d'}^T\boldsymbol{e}_d = \sum_n \boldsymbol{b}_{d'}[n]\boldsymbol{e}_d[n] = 0$, in

the form

$$\boldsymbol{t}_d = (\boldsymbol{t}_d^T \boldsymbol{b}_{d'})\boldsymbol{b}_{d'} + \boldsymbol{e}_d. \tag{C2}$$

Trivially, $\boldsymbol{e}_d = \boldsymbol{t}_d - \mu(\boldsymbol{t}_{d'})\mathbb{1}_{d'}$, where $\mu$ is used to denote the sample average of $\boldsymbol{t}$ over $d'$, $\mu(\boldsymbol{t}_{d'}) = \frac{1}{d'}\sum_{n:|n|\leqslant\frac{d'}{2}} \boldsymbol{t}_{d'}[n]$, and $\mathbb{1}_{d'}$ is a vector which is 1 for $|n| \leqslant \frac{d'}{2}$ and 0 for $|n| > \frac{d'}{2}$. Omitting the covariance normalization for simplicity, the matched filter detector with template $\boldsymbol{b}_{d'}$ is given by

$$T(\boldsymbol{y}; \boldsymbol{b}_{d'}) = \frac{\boldsymbol{y}^T \boldsymbol{b}_{d'}}{\sqrt{\boldsymbol{b}_{d'}^T \boldsymbol{b}_{d'}}} = \boldsymbol{y}^T \boldsymbol{b}_{d'}. \tag{C3}$$

If $\boldsymbol{y} = A\boldsymbol{t}_d + \boldsymbol{n}$,

$$\begin{aligned}\mathbb{E}[T(\boldsymbol{y}; \boldsymbol{b}_{d'})] &= \mathbb{E}[(A\boldsymbol{t}_d + \boldsymbol{n})^T \boldsymbol{b}_{d'}] = A\boldsymbol{t}_d^T \boldsymbol{b}_{d'} \\ &= A((\boldsymbol{t}_d^T \boldsymbol{b}_{d'})\boldsymbol{b}_{d'} + \boldsymbol{e}_d)^T \boldsymbol{b}_d = A(\boldsymbol{t}_d^T \boldsymbol{b}_{d'}).\end{aligned} \tag{C4}$$

Repeating the same calculation with a matched filter using the true transit template (the Neyman–Pearson optimal detector):

$$\mathbb{E}[T(\boldsymbol{y}; \boldsymbol{t}_d)] = A. \tag{C5}$$

Therefore, the loss in detection strength, using a box template of length $d'$ as compared to the true transit template of length $d$, is exactly given by $\boldsymbol{t}_d^T \boldsymbol{b}_{d'}$. This is the same quantity as the fitting factor in B. Allen (2005) and S. Seader et al. (2013, their Section 5), or $1 - \epsilon$, where $\epsilon$ is termed the template mismatch.

In numerical simulations of limb-darkened transit events over the range described in our simulated injection tests in Section 2.3 and box functions of the same $d' = d$, this is approximately $\frac{\mathbb{E}[T(\boldsymbol{y};\boldsymbol{b}_d)]}{\mathbb{E}[T(\boldsymbol{y};\boldsymbol{t}_d)]} \sim 0.922$. When $d'$ is allowed to vary among a discrete set of searched durations in Table 1, the optimal $d'$ that minimizes the loss in detection strength attains $\frac{\mathbb{E}[T(\boldsymbol{y};\boldsymbol{b}_{d'})]}{\mathbb{E}[T(\boldsymbol{y};\boldsymbol{t}_d)]} \sim 0.960$. When $d'$ is allowed to take on any continuous value less than $d$, the optimal $d'$ attains $\frac{\mathbb{E}[T(\boldsymbol{y};\boldsymbol{b}_{d'})]}{\mathbb{E}[T(\boldsymbol{y};\boldsymbol{t}_d)]} \sim 0.968$. This suggests that, on average, a box transit approximation produces a minimal loss in detection power.

S. Seader et al. (2013) also evaluated the fitting factor for a box template with Monte Carlo tests, however in the context of Kepler 30 minutes cadence data. Their tests examined both mismatch due to shape with $d' = d$ and due to parameter discretization with $d' = \arg\min_{d' \in D_{\rm TPS}} \|d' - d\|$, where $D_{\rm TPS}$ is a set of 14 logarithmic spaced durations between 1.5 and 15 hr. They note a loss in fitting factor of 0.0041 (Table 1) for the latter case. In our tests where we vary $d'$, we instead minimize the loss in detection strength as $d' = \arg\max_{d' \in D} \|\boldsymbol{t}_d^T \boldsymbol{b}_{d'}\|$, where $D$ is the discrete set of searched transit durations in Table 1. As noted above, a $d' < d$ is optimal, and for TESS 2 minutes cadence data we observe an improvement in fitting factor of 0.038 as compared to $d' = d$.

## Appendix D
## Transit Search Space Properties

The transit template search space $\boldsymbol{T}$ is defined over the parameter range in transit duration $d$, period $P$, and epoch $e$ defined in Table 1. The search space $\boldsymbol{T}$ can be decomposed over subspace $\boldsymbol{T}_d$ for each discrete duration $d \in D$ as $\boldsymbol{T} = \bigcup_{d \in D} \boldsymbol{T}_d$.

Each duration subspace $\boldsymbol{T}_d$ spans a full sampling of period and epoch. The transit search templates are realized at a discrete sampling increment $\delta P$ in period and $\delta e$ in epoch. If a minimum of three transit events is required for detection of a transit signal in a light curve of length $N$, then the maximum detectable period is $P_{\rm max} = \frac{N}{2}$. Assuming a minimum period $\delta P$, the set of candidate periods is $P \in (\delta P) \cdot \{1, 2, .., \frac{N}{2(\delta P)}\}$. For a given period $P$, the set of candidate epochs is $e \in (\delta e) \cdot \{1, 2, \ldots, \frac{P}{\delta e}\}$.

The number of epochs for period $P$ is $N_e(P) = \frac{P}{\delta e}$. The total number of $(P, e)$ samples in each subspace $\boldsymbol{T}_d$, equivalently the subspace size $|\boldsymbol{T}_d|$, is obtained by summing over the period:

$$|\boldsymbol{T}_d| = \sum_{k=1}^{\frac{N}{2(\delta P)}} N_e(k(\delta P)) = \frac{1}{2(\delta e)}\left(\frac{N^2}{4(\delta P)} + \frac{N}{2}\right). \tag{D1}$$

The size of the full transit search space is

$$|\boldsymbol{T}| = |D| \cdot |\boldsymbol{T}_d| = \frac{|D|}{2(\delta e)}\left(\frac{N^2}{4(\delta P)} + \frac{N}{2}\right) \approx \frac{|D|N^2}{8(\delta e)(\delta P)}. \tag{D2}$$

The number of transit events for fixed period $P$ in a light curve of length $N$ is $\frac{N}{P}$. The total number of transit events $N_{t,d}$ in each transit search subspace $\boldsymbol{T}_d$ is therefore

$$N_{t,d} = \sum_{k=1}^{\frac{N}{2(\delta P)}} \frac{N_e(k(\delta P))N}{k(\delta P)} = \frac{N^2}{2(\delta e)(\delta P)}. \tag{D3}$$

The average number of transit events $\mu_t$ per template in $\boldsymbol{T}_d$ is constant:

$$\mu_t = \frac{N_{t,d}}{|\boldsymbol{T}_d|} \sim 4. \tag{D4}$$

From Table 1, $|D| = 10$ and $\delta P = \delta e = 5\triangle_{\rm LC}$ 2 minutes cadence samples (or 10 minutes); then $|\boldsymbol{T}| \sim \frac{N^2}{20}$. For $N \sim 2 \times 10^5$ (a year of 2 minutes cadence samples) this yields $|\boldsymbol{T}| \leqslant 10^9$. In practice, $|\boldsymbol{T}| \approx 10^8$ as transits that occur on masked data are excluded and a larger step size of 20 minutes is used for candidate periods greater than 27 days (Table 1).

## ORCID iDs

Jamila S. Taaki https://orcid.org/0000-0001-5475-1975
Athol J. Kemball https://orcid.org/0000-0001-6233-8347

## References

Aigrain, S., Parviainen, H., & Pope, B. J. S. 2016, MNRAS, 459, 2408
Allen, B. 2005, PhRvD, 71, 062001
Angus, R., Foreman-Mackey, D., & Johnson, J. A. 2016, ApJ, 818, 109
Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., & Rajpaul, V. 2017, MNRAS, 474, 2094
Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Auvergne, M., Bodin, P., Boisnard, L., et al. 2009, A&A, 506, 411
Barclay, T., Pepper, J., & Quintana, E. V. 2018, ApJS, 239, 2
Battley, M., Kunimoto, M., Armstrong, D., & Pollacco, D. 2021, MNRAS, 503, 4092
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Sci, 327, 977
Burke, C. J., Gaudi, B. S., DePoy, D., & Pogge, R. W. 2006, AJ, 132, 210
Chakraborty, J., Wheeler, A., & Kipping, D. 2020, MNRAS, 499, 4011

Christiansen, J., & Machalek, P. 2010, Kepler Data Release 7 Notes Kepler Science Document KSCI-19047-001, NASA Ames Research Center, archive.stsci.edu/kepler/release_notes/release_notes7/DataRelease_07_2010091618.pdf
Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2013, ApJS, 207, 35
Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2015, ApJ, 810, 95
Christiansen, J. L., Clarke, B. D., Burke, C. J., et al. 2016, ApJ, 828, 99
Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, PASP, 124, 1279
Deeg, H. J., & Alonso, R. 2018, in Handbook of Exoplanets, ed. H. J. Deeg & J. Belmonte (Cham: Springer), 1
Eisner, N. L., BarragÃąn, O., Lintott, C., et al. 2020, MNRAS, 501, 4669
Fausnaugh, M., Morgan, E., Vanderspek, R., et al. 2021, PASP, 133, 095002
Fausnaugh, M. M., Caldwell, D. A., Jenkins, J. M., et al. 2018, TESS Data Release Notes: Sector 1, DR1 Technical Report 20190002535, NASA Ames Research Center, archive.stsci.edu/tess/tess_drn.html
Feinstein, A. D., Montet, B. T., Ansdell, M., et al. 2020, AJ, 160, 219
Foreman-Mackey, D., Montet, B. T., Hogg, D. W., et al. 2015, ApJ, 806, 215
Garcia, L. J., Foreman-Mackey, D., Murray, C. A., et al. 2024, AJ, 167, 284
Gilliland, R. L., Brown, T. M., Guhathakurta, P., et al. 2000, ApJL, 545, L47
Gilliland, R. L., Chaplin, W. J., Dunham, E. W., et al. 2011, ApJS, 197, 6
Guerrero, N. M., Seager, S., Huang, C. X., et al. 2021, ApJS, 254, 39
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Hedges, C., Angus, R., Barentsen, G., et al. 2020, RNAAS, 4, 220
Holdsworth, D. L., Cunha, M. S., Kurtz, D. W., et al. 2021, MNRAS, 506, 1073
Huang, C. X., Vanderburg, A., Pál, A., et al. 2020, RNAAS, 4, 204
Hunter, J. D. 2007, CSE, 9, 90
Jara-Maldonado, M., Alarcon-Aquino, V., Rosas-Romero, R., Starostenko, O., & Ramirez-Cortes, J. M. 2020, EScIn, 13, 573
Jenkins, J. M. 2017, Kepler Data Processing Handbook: Overview of the Science Operations Center Kepler Science Document KSCI-19081-002, NASA Ames Research Center
Jenkins, J. M. 2020, Kepler Data Processing Handbook: Overview of the Science Operations Center, id. 2 Kepler Science Document KSCI-19081-003, NASA Ames Research Center
Jenkins, J. M., Caldwell, D. A., & Borucki, W. J. 2002, ApJ, 564, 495
Jenkins, J. M., Chandrasekaran, H., McCauliff, S. D., et al. 2010, Proc. SPIE, 7740, 140
Jenkins, J. M., Chandrasekaran, H., McCauliff, S. D., et al. 2010, Proc. SPIE, 7740, 77400D
Jenkins, J. M., Doyle, L. R., & Cullers, D. K. 1996, Icar, 119, 244
Jenkins, J. M., Twicken, J. D., McCauliff, S., et al. 2016, Proc. SPIE, 9913, 99133E
Justesen, A. B., & Albrecht, S. 2021, ApJ, 912, 123
Kay, S. 1998, Fundamentals of Statistical Signal Processing: Detection Theory, Vol. 2 (Englewood Cliffs, NJ: Prentice-Hall)
Kay, S. 1999, ITSP, 47, 10
Kay, S. M. 1993, Fundamentals of Statistical Signal Processing: Estimation Theory, Vol. 1 (Englewood Cliffs, NJ: Prentice-Hall)
Kipping, D. M. 2013a, MNRAS, 435, 2152
Kipping, D. M. 2013b, MNRAS, 434, L51
Kovács, G., Bakos, G., & Noyes, R. W. 2005, MNRAS, 356, 557
Kovács, G., Hartman, J. D., & Bakos, G. Á. 2016, A&A, 585, A57
Kunimoto, M., Bryson, S., Daylan, T., et al. 2023, RNAAS, 7, 7
Lu, Y., Angus, R., Agüeros, M. A., et al. 2020, AJ, 160, 168
Lund, M. N., Handberg, R., Buzasi, D. L., et al. 2021, ApJS, 257, 53
Mandel, K., & Agol, E. 2002, ApJL, 580, L171
MAST Team 2021, TESS Light Curves - All Sectors, STScI/MAST, doi:10.17909/t9-nmc8-f686
Matijevič, G., Prša, A., Orosz, J. A., et al. 2012, AJ, 143, 123
Melton, E. J., Feigelson, E. D., Montalto, M., et al. 2024, AJ, 167, 202
Mendes, C. L., Bode, B., Bauer, G. H., et al. 2015, JComS, 10, 327
Mitnyan, T., Borkovits, T., Rappaport, S. A., Pál, A., & Maxted, P. F. L. 2020, MNRAS, 498, 6034
Morris, R. L., Twicken, J. D., Smith, J. C., et al. 2020, Kepler Data Processing Handbook: Photometric Analysis, Kepler Science Document Kepler Science Document KSCI-19081-003, NASA Ames Research Center
Moulin, P., & Veeravalli, V. V. 2018, Statistical Inference for Engineers and Data Scientists (Cambridge: Cambridge Univ. Press)
NExScI 2022, Exoplanet Follow-up Observing Program Web Service, IPAC, doi:10.26134/ExoFOP5
Nguyen, T., Morgan, E., Vanderspek, R., et al. 2018, JATIS, 4, 1
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Prša, A., Kochoska, A., Conroy, K. E., et al. 2022, ApJS, 258, 16
Rasmussen, C. E. 2003, In Summer School on Machine Learning (Berlin: Springer), 63
Ricker, G. R., Vanderspek, R., Winn, J. N., et al. 2016, Proc. SPIE, 9904, 99042B
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014a, JATIS, 1, 1
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014b, JATIS, 1, 014003
Roberts, S., McQuillan, A., Reece, S., & Aigrain, S. 2013, MNRAS, 435, 3639
Santerne, A., Fressin, F., Dìaz, R. F., et al. 2013, A&A, 557, A139
Seader, S., Tenenbaum, P., Jenkins, J. M., & Burke, C. J. 2013, ApJS, 206, 25
Smith, J. C., Stumpe, M. C., Cleve, J. E. V., et al. 2012, PASP, 124, 1000
Stassun, K. G., Oelkers, R. J., Pepper, J., et al. 2018, AJ, 156, 102
Stassun, K. G., Oelkers, R. J., Paegert, M., et al. 2019, AJ, 158, 138
Stumpe, M. C., Smith, J. C., Catanzarite, J. H., et al. 2014, PASP, 126, 100
Stumpe, M. C., Smith, J. C., Cleve, J. E. V., et al. 2012, PASP, 124, 985
Taaki, J. S., Kamalabadi, F., & Kemball, A. J. 2020, AJ, 159, 283
Taaki, J. S., Kemball, A. J., & Kamalabadi, F. 2024, AJ, 167, 60
Tenenbaum, P., & Jenkins, J. M. 2018, TESS Science Data Products Description Document EXP-TESS-ARC-ICD-0014 Rev D, NASA Ames Research Center, ntrs.nasa.gov/api/citations/20180007935/downloads/20180007935.pdf
Thompson, S. E., Coughlin, J. L., Hoffman, K., et al. 2018, ApJS, 235, 38
Twicken, J., Caldwell, D. A., Davies, M. D., et al. 2019, AAS Meeting, 233, 140.03
Twicken, J. D., Catanzarite, J. H., Clarke, B. D., et al. 2018, PASP, 130, 064502
Twicken, J. D., Chandrasekaran, H., Jenkins, J. M., et al. 2010, Proc. SPIE, 7740, 77401U
Twicken, J. D., Clarke, B. D., Bryson, S. T., et al. 2010, Proc. SPIE, 7740, 774023
Twicken, J. D., Jenkins, J. M., Seader, S. E., et al. 2016, AJ, 152, 158
Van Trees, H. L. 2001, Detection, Estimation, and Modulation Theory (New York: Wiley), 239
VanderPlas, J. T. 2018, ApJS, 236, 16
Vanderspek, R., Doty, J., Fausnaugh, M., & Jenkins, J. M. 2018, TESS Instrument Handbook (Baltimore, MD: STScI)
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Weldrake, D. T., Sackett, P. D., Bridges, T. J., & Freeman, K. C. 2005, ApJ, 620, 1043
Wasserman, L. 2013, All of Statistics: A Concise Course in Statistical Inference (New York: Springer),
Wheeler, A., & Kipping, D. 2019, MNRAS, 485, 5498
Wong, I., Shporer, A., Daylan, T., et al. 2020, AJ, 160, 155
Zhu, W., & Dong, S. 2021, ARA&A, 59, 291